Numerical Weather Prediction

The Operational Mesoscale Data Assimilation System 1999-2001: Implementation of 3DVar and Later Upgrades

Forecasting Research Technical Report No. 374

Bruce Macpherson, Phil Andrews, Stephen Harcourt, Bruce Ingleby, Byron Chalcraft, Adam Maycock, Richard Renshaw, Colin Parrett, Simon Anderson, Martin Sharpe, Dawn Harrison and Mark Gibson

email: nwp_publications@metoffice.com

©Crown Copyright

Forecasting Research Technical Report No 374

The Operational Mesoscale Data Assimilation System 1999-2001:

Implementation of 3DVar and Later Upgrades

by

Bruce Macpherson, Phil Andrews, Stephen Harcourt, Bruce Ingleby,
Byron Chalcraft, Adam Maycock, Richard Renshaw,
Colin Parrett, Simon Anderson, Martin Sharpe,
Dawn Harrison, Mark Gibson.

Abstract

The Mesoscale version of 3DVar was introduced operationally in October 1999, replacing the previous Analysis Correction (AC) scheme. Since then, it has undergone significant upgrades in June 2000 and February 2001. This technical report documents the various trials leading to the original implementation and subsequent enhancements. The contents of the upgrade 'packages' are specified, along with subjective and objective assessment of their impact on key forecast variables.

The overall signal from the first version of 3DVar is probably best described as a slight benefit, with the main improvements coming in the analysis and short-period forecast of fog, and prediction of light rain in summertime. The main impact of the June 2000 upgrade, involving a revision of the forecast error covariances, proved to be a slight improvement for precipitation. The February 2001 upgrade brought an improvement in screen temperature resulting from introduction of a soil temperature increment. It also saw the introduction of wind profilers and higher density Meteosat wind data.

Alongside the 3DVar assimilation of most observation types, a nudging scheme is still employed for assimilation of MOPS cloud data from the Nimrod nowcasting system. Also a latent heat nudging scheme is used to assimilate rain rate estimates derived from radar data. Developments in these schemes during this period are also described. A change to the quality control of MOPS cloud data was made in March 2000, reducing the analysed cloud bias. A radar quality estimate was introduced into the latent heat nudging scheme in September 2000. European radar data have extended the coverage available for assimilation since February 2001.

CONTENTS

page

1. Implementation of 3DVar in the UK Mesoscale Model…..……....3

2. The June 2000 Upgrade to 3DVar……..…..……...26

3. The February 2001 Upgrade to 3DVar…….…….……....39

4. Developments in cloud assimilation…….…….52

5. Developments in rainfall assimilation……...59

6. References…….64

1. Implementation of 3DVar in the UK Mesoscale Model

1.1 Specification of Mes 3DVar

The global version of 3DVar was implemented operationally in March 1999 and has been described by Lorenc et al. (2000). The Mesoscale (Mes) version of 3DVar is similar in most respects to its global counterpart, but there are a few differences worth noting:

- a constraint which forces analysis increments to be zero at the boundary for scalar fields and the normal component of the wind. This is achieved via a double-sine transform.
- a visibility analysis which gives increments to aerosol and moisture fields. The visibility observation operator is nonlinear, and the minimisation algorithm used by Mes 3DVar is appropriate to nonlinear operators. A full treatment of the visibility analysis will be given in a companion paper (Harcourt and Macpherson, *in preparation*), but the multivariate 3DVar analysis is superior in formulation to its predecessor, the Analysis Correction (AC) Scheme, in which visibility observations could only give rise to aerosol increments.
- forecast error covariances based on 100 days of (t+12-t+6) forecast differences for deriving vertical modes. For a description of the forecast error representation in global 3DVar, see Ingleby (2001).
- a fixed scale of 90km in the SOAR (2^{nd} Order Autoregressive) function used in the horizontal analysis of all variables.
- a nudging (AC) scheme for the assimilation of MOPS cloud and rainfall data, alongside the 3DVar scheme for analysis of all other data sources. For background on these schemes for cloud and rainfall assimilation, see Macpherson et al. (1996), Appendix A of Radcliffe et al. (1998) and Jones and Macpherson (1997). The MOPS data are nudged into the model during the Incremental Analysis Update (IAU) run, when 3DVar increments are added gradually in a two-hour window centred on t+0.
- Mes 3DVar makes greater use of SYNOP data than the global model, but its first implementation made reduced use of hourly SYNOPs relative to the Mes AC scheme, with only the nearest Synop to t+0 selected instead of data from each hour.

1.2 The parallel trial

The 'real-time' parallel trial ran in two phases from 7th-22nd June 1999 and from 25th June-17th July. The trial was coupled interactively to the Nimrod nowcasting system, so that forecast output from the trial run was fed to a specially run trial version of Nimrod to generate the cloud and precipitation data for the trial. This was necessary in view of uncertainty about the impact of 3DVar on the short period forecast used by Nimrod.

The trial used identical boundary conditions to the operational run as supplied by the 'prelim' run of the global forecast.

1.3 Case studies

Ten cases were chosen mainly from a set used previously for the validation of MOSES (the land surface scheme) and the latest boundary layer scheme (Radcliffe et al., 2001). A couple of particularly foggy cases were included to test the new visibility analysis more thoroughly. The case studies were run with a version of the model equivalent to that running operationally in June 1999. Each included 12 hours of assimilation followed by a forecast to t+36. The start field for assimilation included soil moisture set from MORECS data in most cases, with climatological soil moisture used for technical reasons in a couple of winter cases. A few differences from the real-time trial should be noted:

- the boundary conditions were taken from the archived global forecast at the time of the case. This was a forecast from a global AC assimilation run, whereas the real-time trial was driven by boundary data from a global 3DVar IAU run.
- the cases were run with a 4-point boundary relaxation zone, current at the time, while the real-time trial used the operational 8-point boundary zone.
- the cases used archived MOPS cloud data, as for technical reasons it is not practical to couple case studies interactively with Nimrod.
- the cases used 3-hourly, rather than hourly, radar data, since hourly data were only introduced in May 1999.

1.4 Subjective assessment

1.4.1 Case Studies

Fields studied were precipitation, PMSL, fog, stratus and (where appropriate) snow probabilities, 1.5 metre temperature, high medium and low cloud cover from both control (pseudo-operational) and 3DVar trial runs. These were then compared against plotted NW Europe surface charts, and where necessary, radar rainfall and satellite imagery. Runs were compared out to a full forecast range of T+36 hours. Only those parameters where there were significant differences between the control and the 3DVar trial run are discussed below.

<u>5th August 1998 (clear summer day)</u>

Fog and stratus probabilities were both a little lower over parts of Normandy but not sufficiently different for there to be a preference.

<u>2nd October 1998 (clear, cold night)</u>

The 3DVar run had somewhat more medium cloud over England from T+6 to T+18 which was preferred. Temperatures were also preferred in the 3DVar run for being 1-2 degrees higher over parts of England on later (night-time) frames.

<u>16th October 1998 (active fronts)</u>

Significant differences were only seen in precipitation, and only at T+9 and T+12. There was no preference at T+9, but at T+12 the 3DVar run was preferred for having less rain over southeast

England. This is shown in Figure 1.1, along with the verifying radar rates.

24th October 1998 (land gales)

Only PMSL showed any significant differences. At T+0 the 3DVar run's PMSL was around 2hPa lower than the control over much of England, with the control preferred. The 3DVar run PMSL was also more 'noisy' in that edge effects propagated more strongly into the domain than in the control, giving a sine wave appearance to the isobars at T+0. This was much reduced by T+3, as shown in Figure 1.2. By T+6 the forecast fields were virtually identical.

26th October 1998 (convection)

No significant differences.

9th November 1998 (organized convection)

There were some differences in fog probability but only at T+0 and T+3. The T+0 frame of the 3DVar run had a much lower probability over part of the central North Sea which looked correct, considering there were 30 knot surface winds in the area. The T+3 frame had a small area of very high probability just over and off of the Dutch coast which also looked better since fog was occurring over Holland and which would have been being blown northwards over the sea.

2nd December 1998 (cold easterlies)

Stratus probability was higher in places on the trial from T+0 to T+15, especially over parts of Ireland, which was preferred. From T+3 to T+9, 1.5 metre temperatures were around 1 degree Celsius lower in the 3DVar run over parts of the east and south of England. The control run verified better for this. There were also some significant differences in this case in PMSL. The 3DVar run was 2-4 hPa lower than the control over the U.K. at T+0, with the control preferred. At T+6 and T+9 the 3DVar run was still 1-2 hPa too low over northern England, but thereafter no significant difference.

22nd December 1998 (ridge, with fog all day in places)

For precipitation the 3DVar run was preferred at T+6, T+9 and T+12 for correctly having less precipitation over southern England. Fog probability was also preferred with 3DVar from T+0 through to T+9 in that it had both a higher probability and a better placement of the fog area. Figure 1.3(a) shows the fog probability output from the T+6 frames. The actual fog area at the time covered a large area as seen in Figure 1.3(b).

12th January 1999 (mixed snow and rain)

Generally neutral. Fog probability was a little lower on the 3DVar run over much of England at T+0 and T+3. The control run was preferred in this respect. Temperatures in the 3DVar run at T+0 *only* were a little lower than the control and were preferred.

21st January 1999 (persistent fog over parts of England)

There were some differences seen in both fog and stratus probability. The 3DVar run was preferred for both, in having a slightly *higher* fog probability and also a higher stratus probability across southern England throughout.

Summary of case studies

The differences were generally fairly small and/or short-lived and for four of the cases there was no overall preference. Of the remaining six cases the 3DVar run was slightly preferred in five and the control run slightly preferred in just one.

It is worth commenting on the pmsl noise problem early in the 3DVar forecast, noticed particularly in the 24th October case, but not seen in the real-time trial. This is probably due to the fact that the case studies were driven by boundary data from a global AC run, while the real-time phase was driven by a global 3DVar 'IAU' run. The global AC run attempts to fit the synoptic hour pressure observations best by t+0, while the IAU run has added only half its observation increments by t+0 and is therefore less likely to generate a discrepancy with the interior mesoscale field. A second difference in the case studies, namely the narrower 4-point boundary zone, may have added to the problem. Since Mes 3DVar is driven operationally by a global IAU run, this problem was not found after implementation.

1.4.2 Subjective assessment of parallel trial

The emphasis of the assessment was on precipitation. However, temperature, stratus and fog probability, PMSL and cloud cover were also assessed for many runs. The periods covered were: 7th to 22nd June, 25th to 26th June, 29th June to 2nd July, 7th to 9th July, 12th to 16th July, a total of 29 days. With the generally neutral impact and the large number of runs, not every available pair of runs was assessed, but at least one pair from each available day was, i.e. 54 pairs in total.

Synoptic types

The period 7th to 22nd June began rather cyclonic and unsettled. High pressure then became rather more dominant, but still with an upper vortex over or close to the U.K. keeping it cool, often cloudy and rather showery until the 13th. This was followed by a warmer ridge for a few days for most parts and then a more cyclonic period again from the 19th to the 21st June. A warm south to southeasterly on the 25th June gave way to thundery troughs moving from the southwest on the 26th. The 29th June to 2nd July was mainly unsettled, with Atlantic fronts crossing the country from time to time. The 7th to the 9th July was anticyclonic, and then after the decline of a high near the U.K. on the 12th there was a generally unsettled westerly type again to the 16th.

Precipitation

In the majority of runs examined (40) it was found that there were either no significant differences at all or that any slight improvements on some forecast frames were balanced by slight detriments at other times. There were a total of 13 runs where the overall 3DVar

precipitation forecast was preferred to the control. These were mostly short-lived or only over a fairly small area. In only one case was the control run's precipitation forecast preferred. This occurred at T+18 and T+21 into the forecast when the 3DVar run had too much rain retained on a weak front over East Anglia on the 25th June and is shown in Figure 1.4. Two examples of improved precipitation forecasts in the 3DVar runs are shown in Figure 1.5 and in Figure 1.6. The latter was the case when play at Wimbledon was abandoned for the day because a slow-moving cold front kept rain over the southeast throughout the afternoon and early evening.

Pressure at mean sea level

There were no significant differences on the majority of runs. On a few occasions low pressure centres were analysed a little (1-2hPa) lower in the 3DVar runs which was preferred.

Cloud cover

No significant differences for the most part. A couple of runs had slightly improved low cloud cover in the trial.

Fog probability

With little fog in the parallel trial period there were few if any significant differences for the majority of runs, but when fog *was* occurring the 3DVar run usually had a slightly better fog analysis, which then fed through to small improvements in the forecast, typically out to T+9. The number of runs where the 3DVar fog forecast was preferred was 8 and where the control was preferred only 2.

Stratus probability

A generally neutral impact once again, but there were 3 runs in which the 3DVar run's forecast was slightly preferred.

1.5 metre temperature

3 runs showed a slight (1-2 degree.) improvement in the temperature forecast over small areas in the 3DVar run for a part of the forecast period, but this appeared to usually be as a result of some small improvements in low cloud cover. It was also noticed that on any clear, relatively cold radiation nights the 3DVar run produced a smoother temperature analysis field.

1.5. Objective Verification

1.5.1 Case Studies

The variables making up the 5-component UK index were verified for the eight 00z and two 06z cases and the results presented are averaged over forecast times from t+6 to t+24 as used in the index, using a station list of 'all UK stations'. The results are summarised in Table 1.1.

Variable	Control		Trial		Change in Skill/ETS
	rms	ETS	rms	ETS	
10-metre wind	3.22		3.23		0
screen temperature	1.35		1.37		-0.01
precipitation		0.46		0.47	+0.01
cloud amount		0.27		0.26 }	
				}	-0.01
cloud base		0.18		0.18 }	
visibility		0.14		0.175	+0.035
		balance of change in skill/ETS			+0.025

Table 1.1 - Summary of UK index components for the case studies

Skill changes for *10-metre wind and screen temperature* were estimated on the basis of mean skill being 0.8 for wind and 0.6 for temperature, with the further assumption that the persistence errors for a t+0 analysis are equal for trial and control. Graphs of the rms errors for wind and temperature are shown in Figure 1.7 and Figure 1.8 respectively.

For *precipitation*, the verification scores are for t+12 and t+24 for 8 cases with DT 00z, since t+6 and t+18 scores are not available (observed 6-hourly accumulations are only available to the verification system at 00z and 12z). For the 2 cases with DT 06z, t+12 and t+24 scores were not available for the same reason, but these 2 cases contributed only a small proportion of the total rainfall events. We note that the subjective assessment of rainfall, favouring 3DVar in 2 cases and the control in none, is consistent with the marginal benefit in the objective scores.

For *cloud*, the change in Equitable Threat Score (ETS) derives from the mean of scores for cloud amount and cloud base, as used in the index.

For *fog*, the results are averaged over both 00z and 6z cases, with the weighting of the average respecting the event frequency of the different thresholds. There were more observations of dense fog (vis<200m) in the 2 'foggy' cases than in the 8 other cases. The improvements noted subjectively in these 2 cases give an important contribution to the ETS improvement in Table 1.1, which comes from the 200m and 1000m thresholds, the 5000m threshold returning an impact close to neutral.

Although not part of the UK index, *pressure at mean sea level* was also verified, with the result shown in Figure 1.9. The trial verifies worse at t+6, which would appear to be a hangover from the boundary noise issue discussed in section 1.4.1. Thereafter, however, the trial verifies better, the balance being a 2% improvement over the period t+6-t+36.

Summary

The balance of the skill/ETS changes is +0.025. Assuming the mean skill to be 0.5, these figures imply a 5% improvement in UK index. This cannot be taken as representative of an annual figure, however, because for example, the frequency of very foggy cases here (2/10) is greater than will occur over a long period. It is more reasonable to say that on the basis of the cases examined, 3DVar is providing a match for the control, and a noticeable benefit for visibility forecasts in very foggy cases.

1.5.2 Parallel trial

The variables making up the 5-component UK index were verified for all available cases over the two phases of the trial from 7th-22nd June and from 25th June-17th July. Results presented are averaged over forecast times from t+6 to t+24 as used in the index, using a station list of 'all UK stations'. The results are summarised in Table 1.2.

Variable	Control		Trial		Change in Skill/ETS
	skill	ETS	skill	ETS	
10-metre wind	0.584		0.592		+0.008
screen temperature	0.494		0.511		+0.017
precipitation		0.262	0.267	+0.007	
cloud amount		0.190	0.187 }		
				}	-0.004
cloud base		0.206	0.201 }		
visibility		0.024	0.031	+0.007	
			balance of change in skill/ETS		+0.035
			sum of control skill components		1.562
			%age change in UK index for trial		+2%

Table 1.2 - Summary of UK index components for the parallel trial

For *10-metre wind and screen temperature*, skill score estimates were available from the continuous trial which are not available from individual case studies. Originally these scores were computed with operational and 3DVar runs each using their own 't+0 persistence' rms error, but the t+0 field supplied from the trial was half way through the IAU period and not therefore the best estimate of a trial analysis for this purpose. For this reason, the skill results for the trial were recomputed using the persistence rms from the operational run. For this period, this made only a small difference to the results. Graphs of mean and rms errors for 10-metre wind and screen temperature for phase 2 of the trial are shown in Figure 1.10 and Figure 1.11.

For *precipitation*, verification was derived from comparison of the model accumulations with those from Nimrod analyses. This form of verification was used for a number of previous mesoscale trials and benefits from the greater representivity of the radar coverage compared to the surface observing network. It is also more consistent with the usual method of subjective assessment.

ETS results are shown in Figure 1.12. The 2mm threshold is not part of the UK index, but is included because it attracts more events than the 4mm threshold and is useful for giving more representative results. It is noticeable that 3DVar is consistently better at the lightest threshold, but starts slightly worse at t+6 for higher thresholds. By t+18 and t+24, 3DVar verifies better at all thresholds, and the larger part of the overall improvement to the index contribution comes from these forecast times. This overall signal is consistent with the balance of comments made subjectively.

The detriment at t+6 for all but light rain may originate in some kind of spin-up process arising from the interaction of the IAU nudging of conventional data increments and the MOPS nudging of cloud and rain data, but this is speculative. The improvements in rainfall later in the forecast would appear to come from 3DVar's better analysis of conventional data.

In Figure 1.13, we plot the Frequency Bias for precipitation events at different thresholds. For this period, 3DVar reduces (improves) the bias for all thresholds. For light rain, it particularly reduces over-prediction early in the forecast. For the heaviest threshold, the reduction is more marked further on in the forecast.

The *cloud* scores show only a small shift in the trial, and we recall that very little difference was seen subjectively.

Although there was negligible dense fog in the real-time trial, the objective *visibility* results at the 1000m and 5000m thresholds favoured the trial, giving the positive increment to the UK index component in Table 1.2, and consistent with the comments made during subjective assessment.

Pressure at mean sea level verified slightly worse for 3DVar, but with the difference between trial and control decaying throughout the forecast, as seen in Figure 1.14 for phase 1 of the trial. From the subjective assessment, which noted no major impacts, it may be concluded that these small objective differences were not associated with any significant differences in evolution.

Summary

The balance of objective verification for the parallel trial favours 3DVar and implies a 2% improvement in the UK index. All variables except cloud showed an improvement.

1.6 Summary of 3DVar implementation trial

A trial of 3DVar in the UK Mesoscale model was completed, involving 5½ weeks of 'real-time' parallel running alongside the operational system in June/July 1999 and 10 case studies drawn from other months.

A subjective assessment of the *'real-time' phase* slightly favoured 3DVar. This is consistent with overall objective verification suggesting a 2% increase in the estimate of the 5-component UK NWP index for 3DVar. A subjective assessment of the *case studies* also favoured 3DVar, with 5/10 cases preferred from the trial and only 1/10 from the control, though the differences were slight in most instances. The objective verification of the case studies showed a 5-component UK index that is 5% better in the trial. One should not extrapolate the index change for this sample of cases into an annual mean figure, but on the basis of the 10 cases examined, 3DVar provides a match for the control, and a significant benefit for visibility forecasts in very foggy cases.

Taking the real-time and case study phases together, the overall signal from 3DVar is probably best described as a slight benefit, with the main improvements in the analysis and short-period forecast of fog, and prediction of light rain in summertime. More specific comments by variable are worth noting:

- **precipitation** in the parallel trial, 3DVar was preferred in 13/54 runs and the AC control in 1/54. Objective verification of 6-hourly accumulations v Nimrod for the real-time trial showed 3DVar slightly better overall, especially for light rain and after t+12 for all thresholds. The frequency bias was reduced (improved) at all thresholds by 3DVar. In subjective assessment of the case studies, 3DVar was preferred in 2/10 cases and the AC control in none, while the objective scores showed a very marginal benefit for 3DVar.
- **visibility** in subjective assessment of the real-time trial, fog probabilities from the 3DVar run were preferred 8 times, and from the operational run 2 times. Although there was no dense fog in the real-time trial, the objective results at the 1000m and 5000m thresholds favoured the trial. In the 2 winter case studies specially chosen to examine impact on dense fog, the trial verified significantly better at the 200m and 1000m thresholds. These improvements were noted subjectively. In the other 8 cases, the trial also performed better at the 1000m threshold.
- **temperature** screen temperature rms errors were lower in the 3DVar runs from the real-time trial over the period t+6-t+24, leading to an estimated skill improvement in 3DVar of 3%. In the 10 case studies, the trial was marginally worse than the control, implying a skill detriment of 2% from 3DVar.
- **wind** 10-metre wind field shows negligible impact from 3DVar.
- **cloud** In subjective assessment, 3/10 cases gave rise to favourable comments on cloud/stratus from the 3DVar run, and in the real-time trial 5/54 3DVar runs were favoured, with no control/operational runs preferred. Differences were, however, only slight and the subjective signal is close to neutral. Objective verification of cloud amount and cloud base showed a very marginal detriment in ETS.
- **pmsl** In the real-time phase, the 3DVar run verified slightly worse objectively (2% over t+6-t+24), but there was no signal detected subjectively. Objective verification showed the 3DVar run worse at t+6, but better thereafter, the balance being a 2% improvement over the period t+6-t+36.

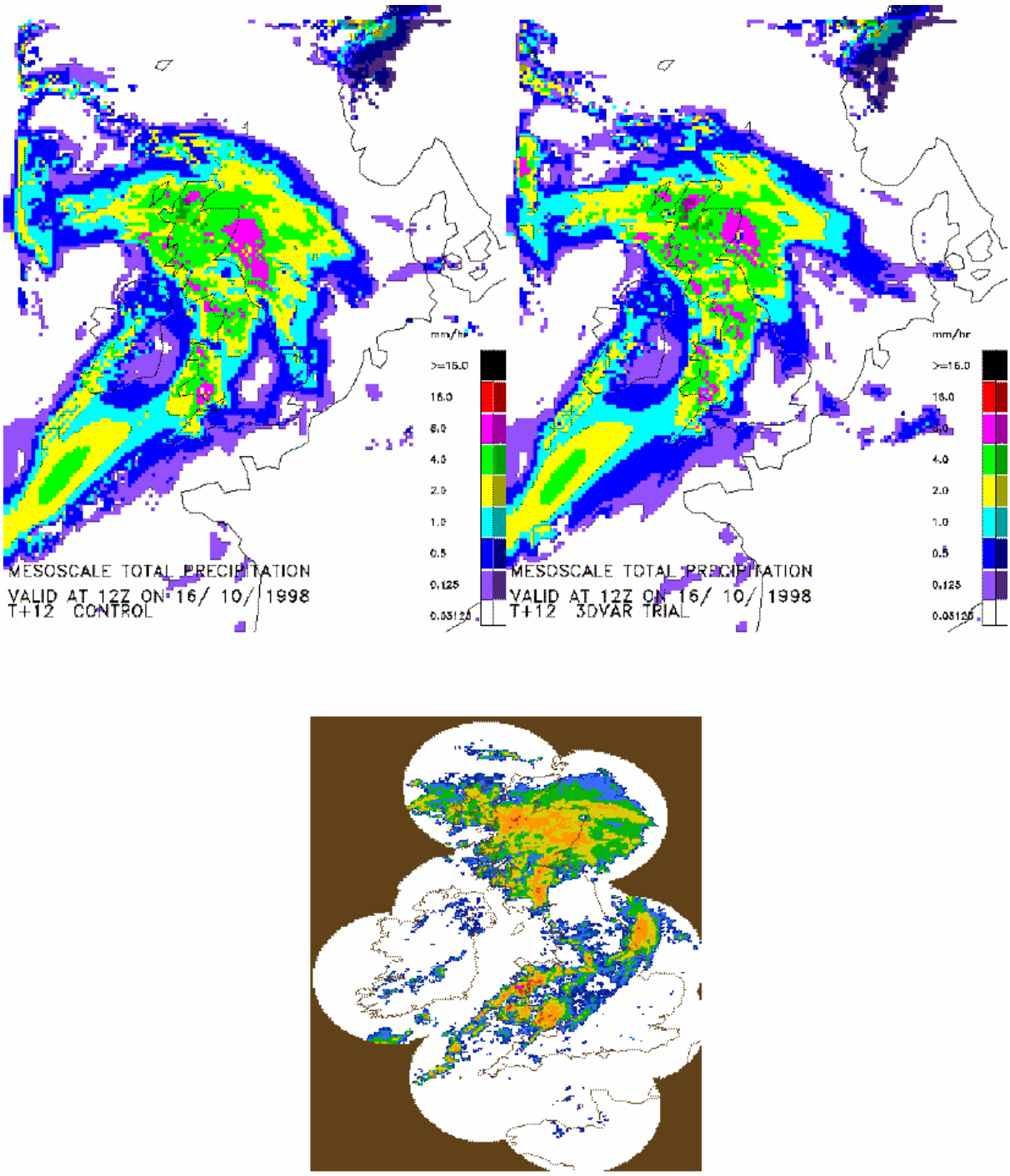

Figure 1.1: 16th October 1998 case study from 3DVar implementation trial.
3DVar rain rate forecast at t+12 (top right) preferred to control forecast at t+12 (top left).
Verifying radar composite in lower frame.

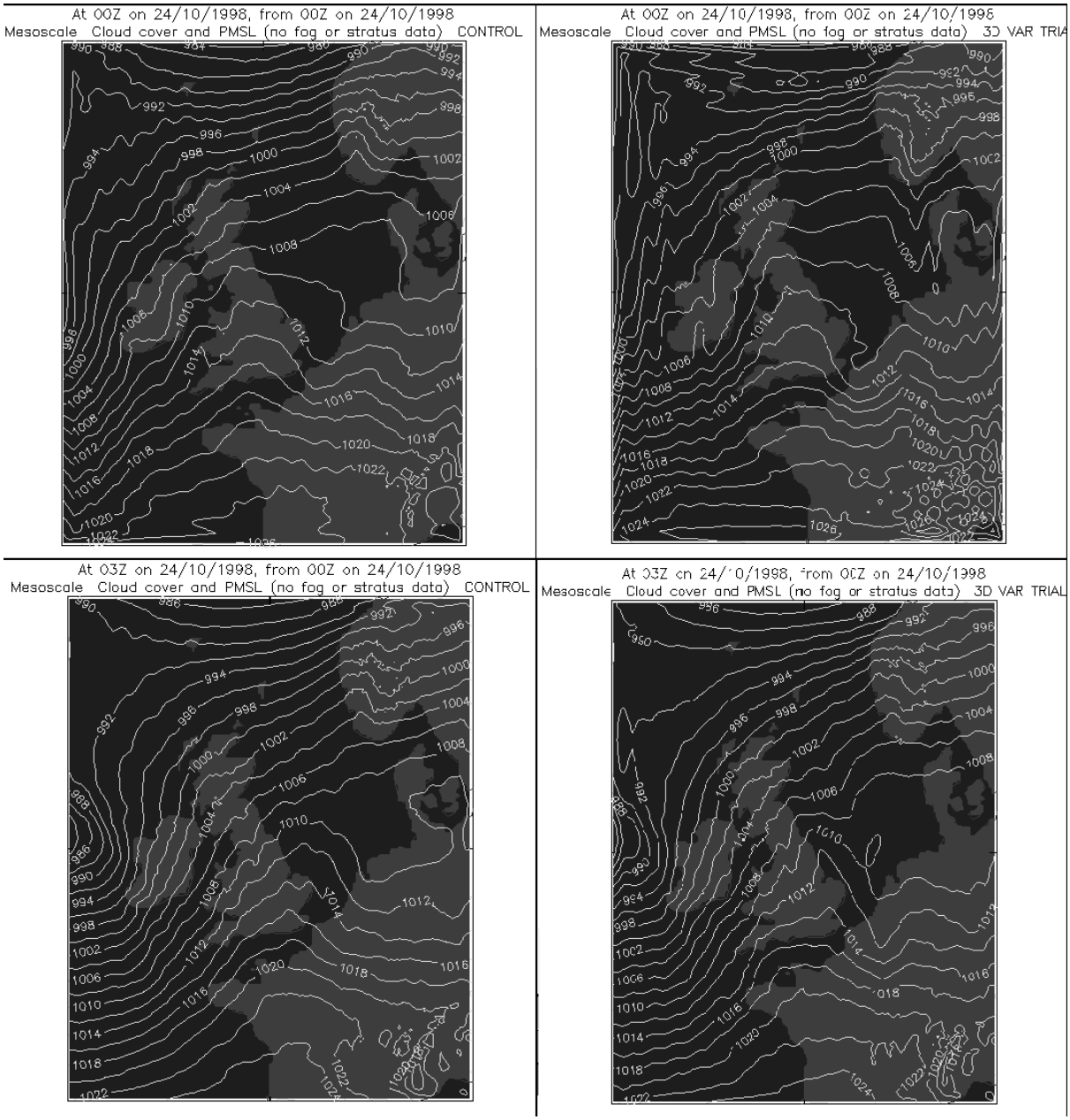

Figure 1.2: 24th October 1998 case study from 3DVar implementation trial.
Control PMSL forecasts on the left, 3DVar forecasts on the right. T+0 frames (top) and t+3 frames (bottom).
The 3DVar t+0 field is noisier than control, but the difference has reduced by t+3.

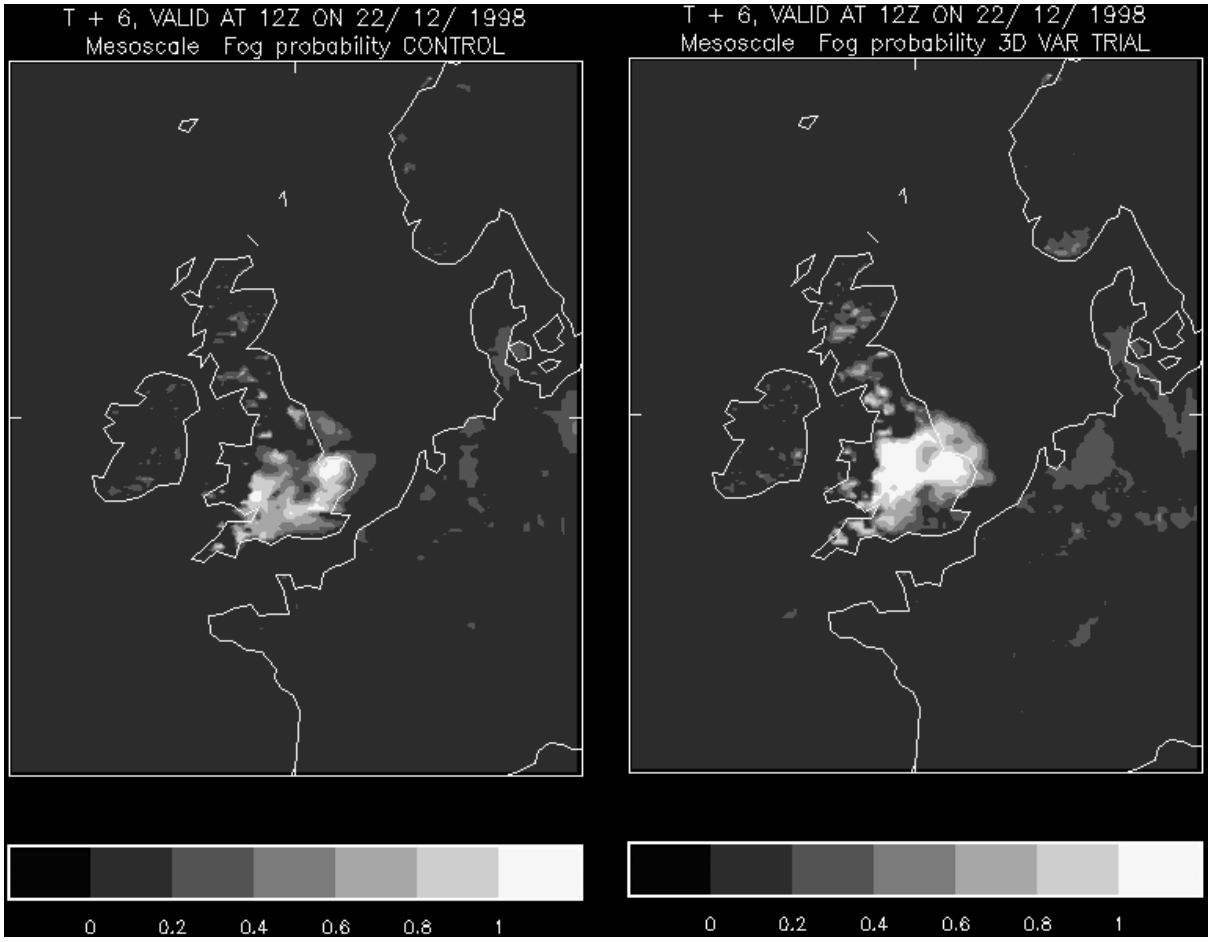

Figure 1.3(a): 22nd December 1998 case study from 3DVar implementation trial.
Fog probability forecasts at t+6 for control (left) and 3DVar (right).
Verifying surface reports of fog are in Figure 1.3(b).
The 3DVar run gives higher (better) probabilities and better placement of the fog area.

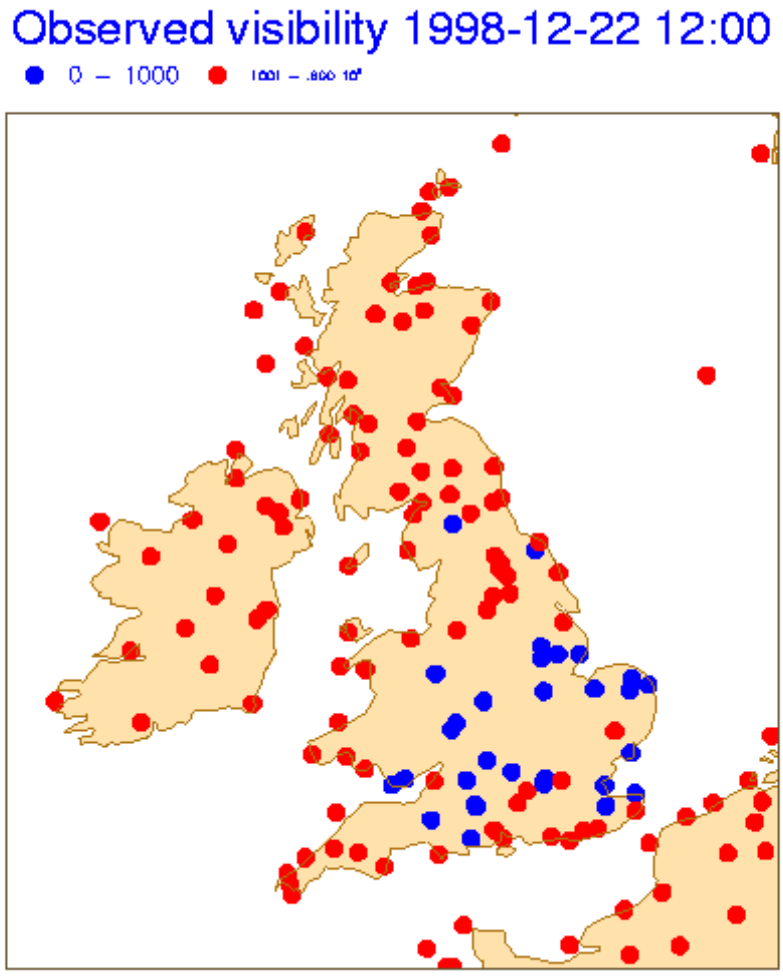

Figure 1.3(b): Reports of visibility below 1000m (blue circles) and above 1000m (red circles) at 12 UTC on 22nd December 1998. See Figure 1.3(a).

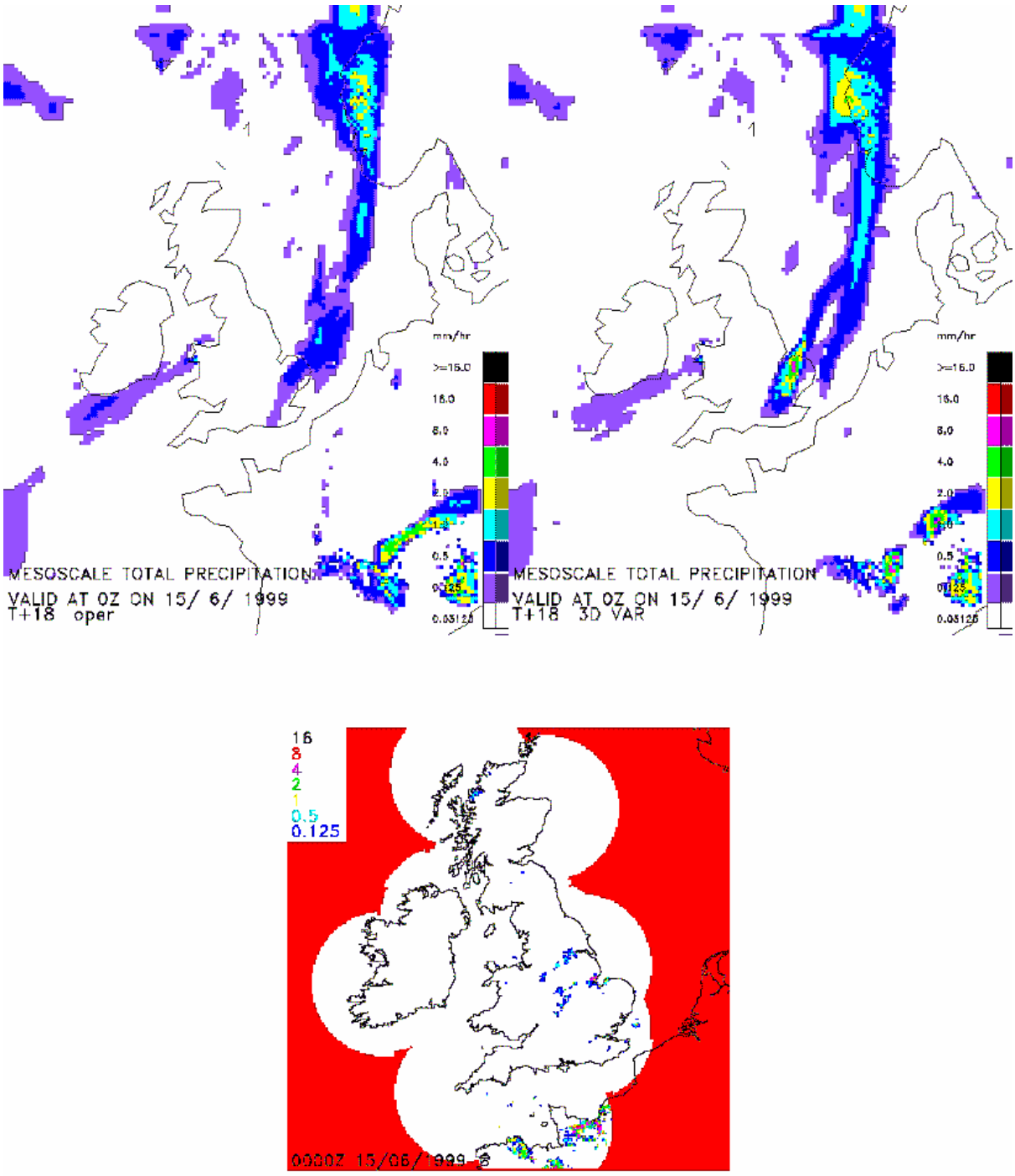

Figure 1.4: Real-time trial of 3DVar implementation. Operational (top left) and 3DVar (top right) forecasts of rain rate at t+18, valid at 0 UTC, 15/6/1999. Verifying radar composite in lower frame. 3DVar run has heavier rain over East Anglia. Control run preferred.

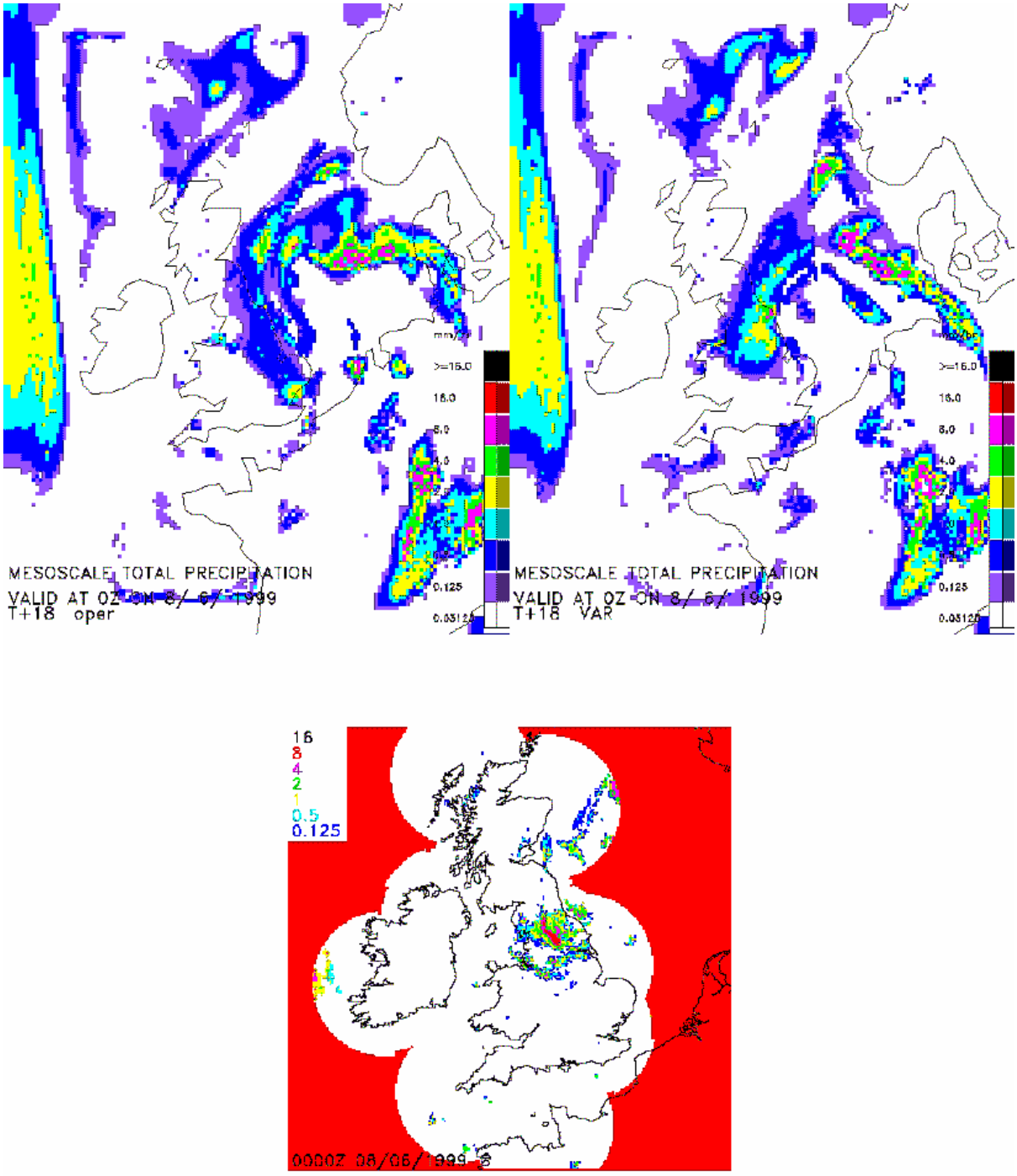

Figure 1.5: Real-time trial of 3DVar implementation. Operational (top left) and 3DVar (top right) forecasts of rain rate at t+18, valid at 0 UTC, 8/6/1999. Verifying radar composite in lower frame. 3DVar run preferred.

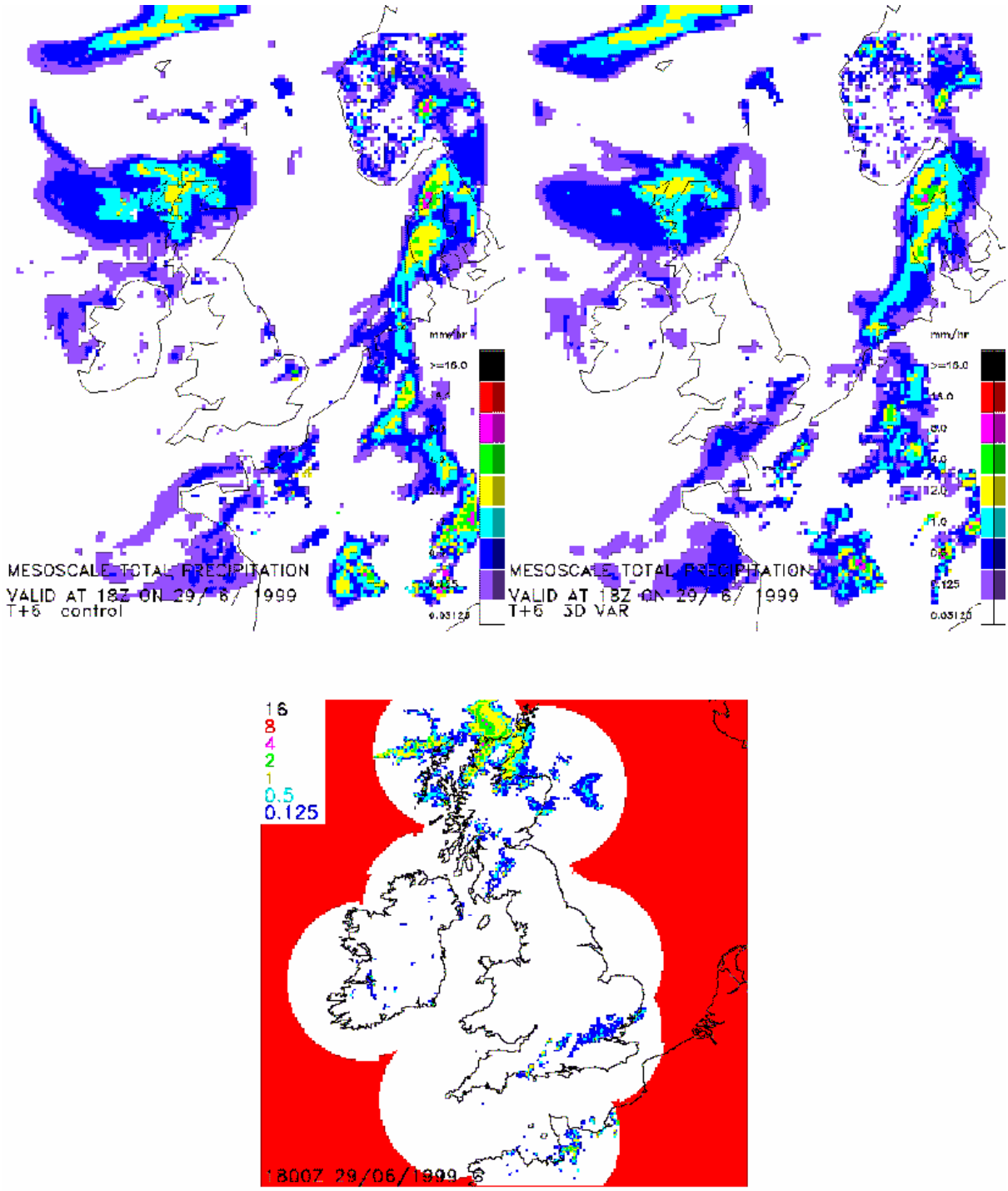

Figure 1.6: Real-time trial of 3DVar implementation. Operational (top left) and 3DVar (top right) forecasts of rain rate at t+6, valid at 18 UTC, 29/6/1999. Verifying radar composite in lower frame. 3DVar run preferred.

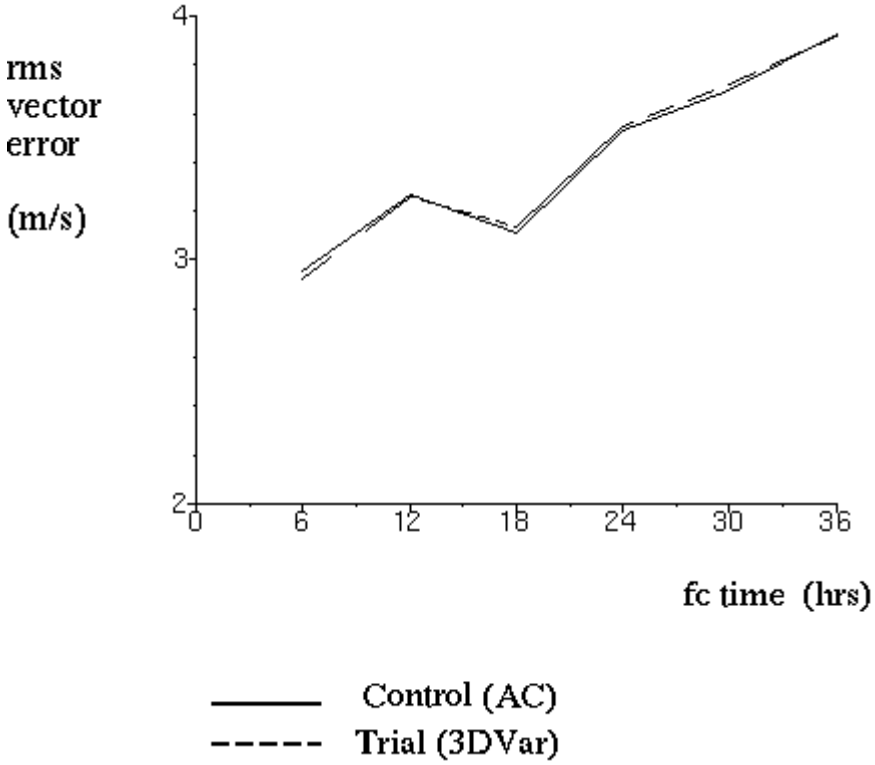

Figure 1.7: Verification of 10-metre wind for 10 cases studies in 3DVar implementation trial. Results averaged over all UK stations.

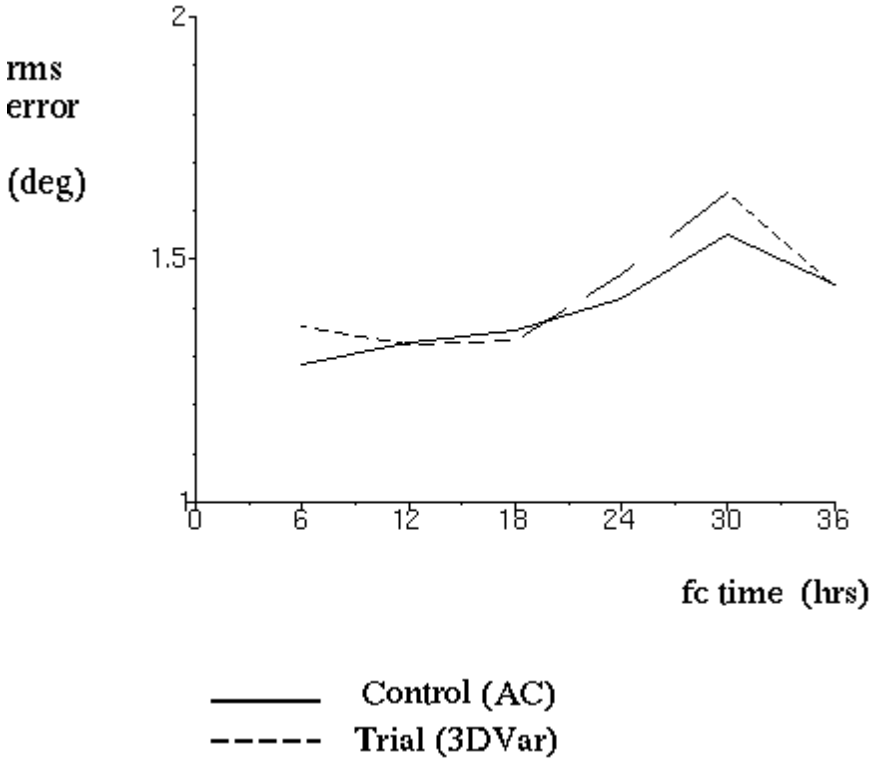

Figure 1.8: Verification of screen temperature for 10 cases studies in 3DVar implementation trial. Results averaged over all UK stations.

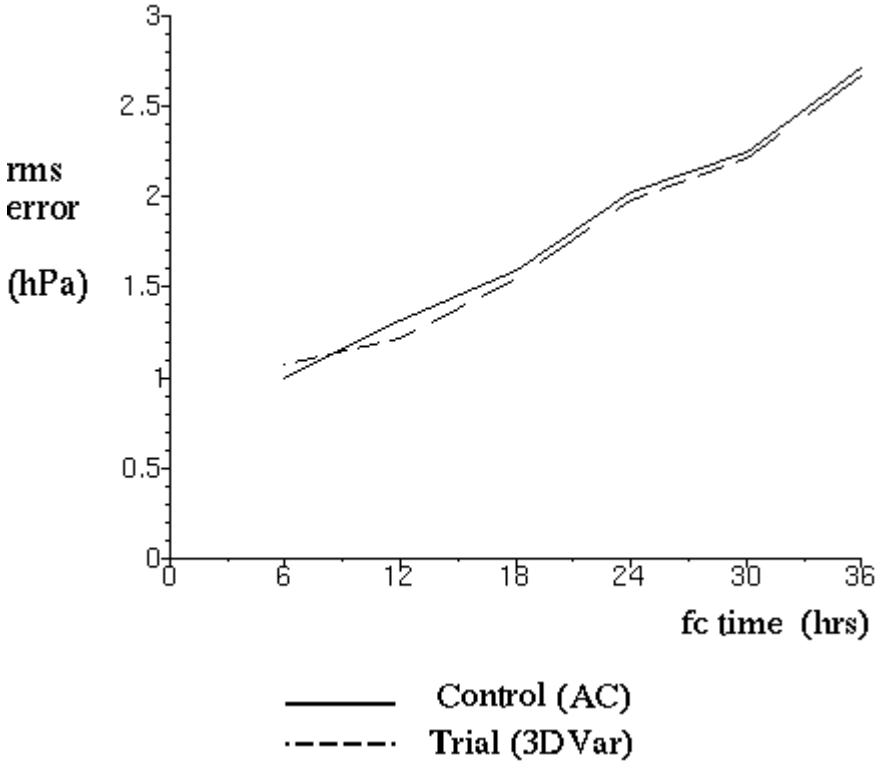

Figure 1.9: Verification of PMSL (hPa) for 10 cases studies in 3DVar implementation trial. Results averaged over all UK stations.

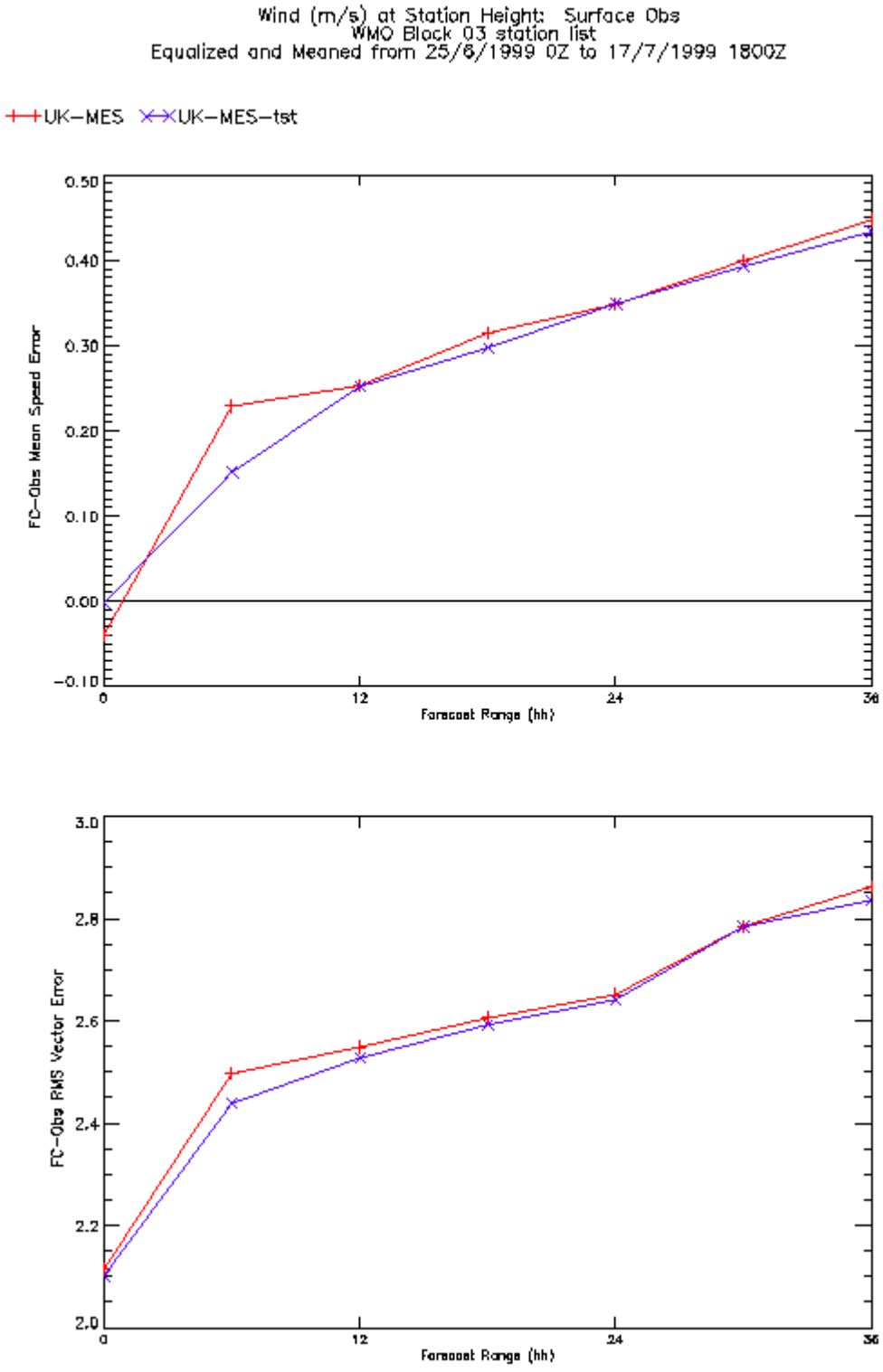

Figure 1.10: Verification of 10-metre wind for real-time phase of 3DVar implementation trial. Results averaged over all UK stations. Control in red, 3DVar in blue. Top frame shows mean speed error, lower frame shows vector wind error.

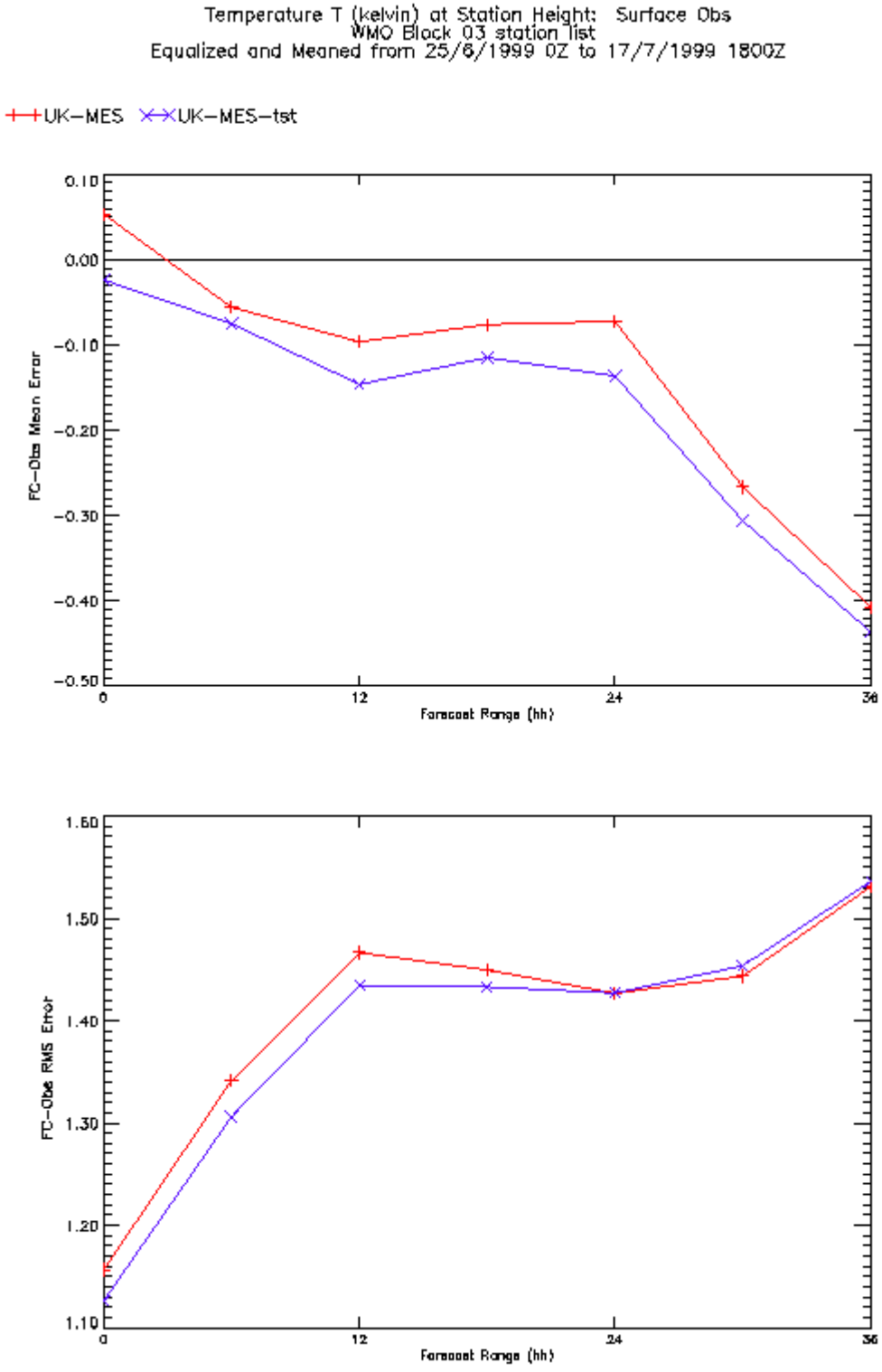

Figure 1.11: Verification of screen temperature for real-time phase of 3DVar implementation trial. Results averaged over all UK stations. Control in red, 3DVar in blue. Top frame shows mean error, lower frame shows rms error.

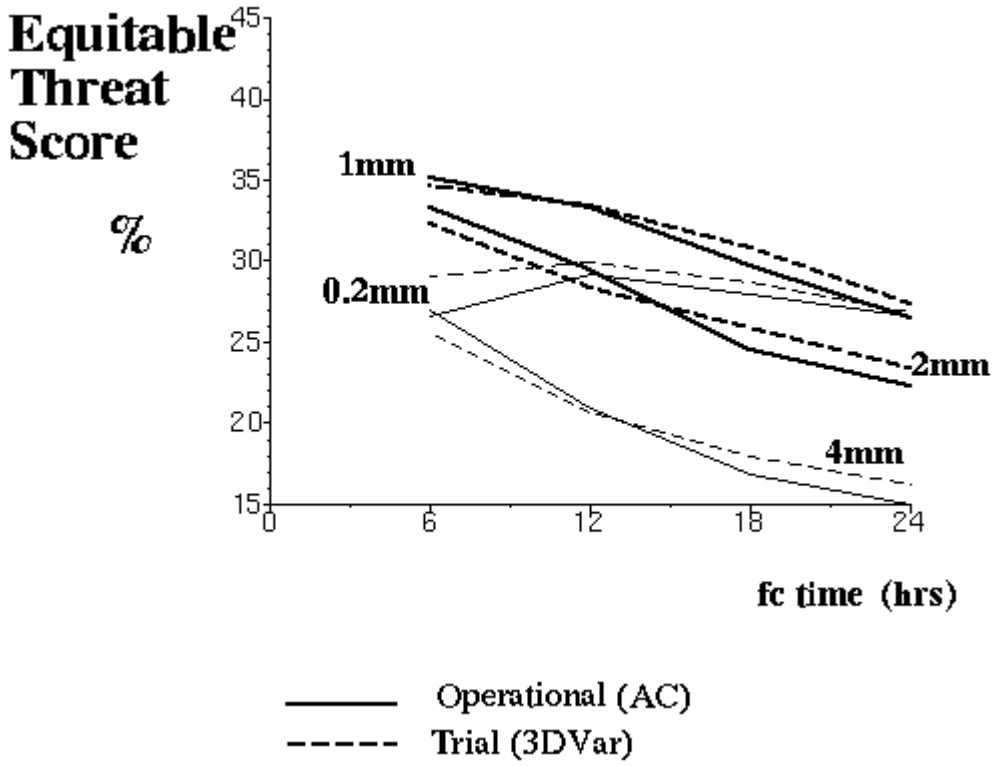

Figure 1.12: Verification of 6-hourly rainfall accumulations against Nimrod (radar) data, for real-time phase of 3DVar implementation trial. ETS results averaged over 94 forecasts from 7th June - 19th July 1999, and presented for 4 different rain amount thresholds as labelled.

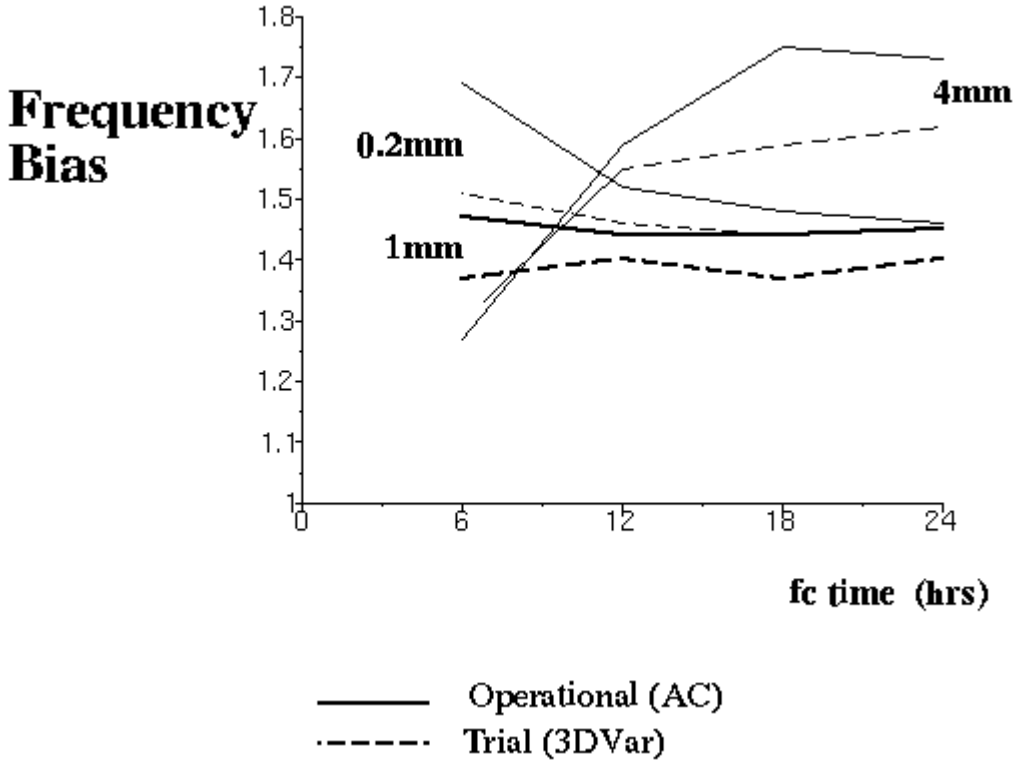

Figure 1.13: As Figure 1.12 but showing the Frequency Bias statistic, the ratio of forecast occurrences to observed occurrences of rainfall amounts of various thresholds.

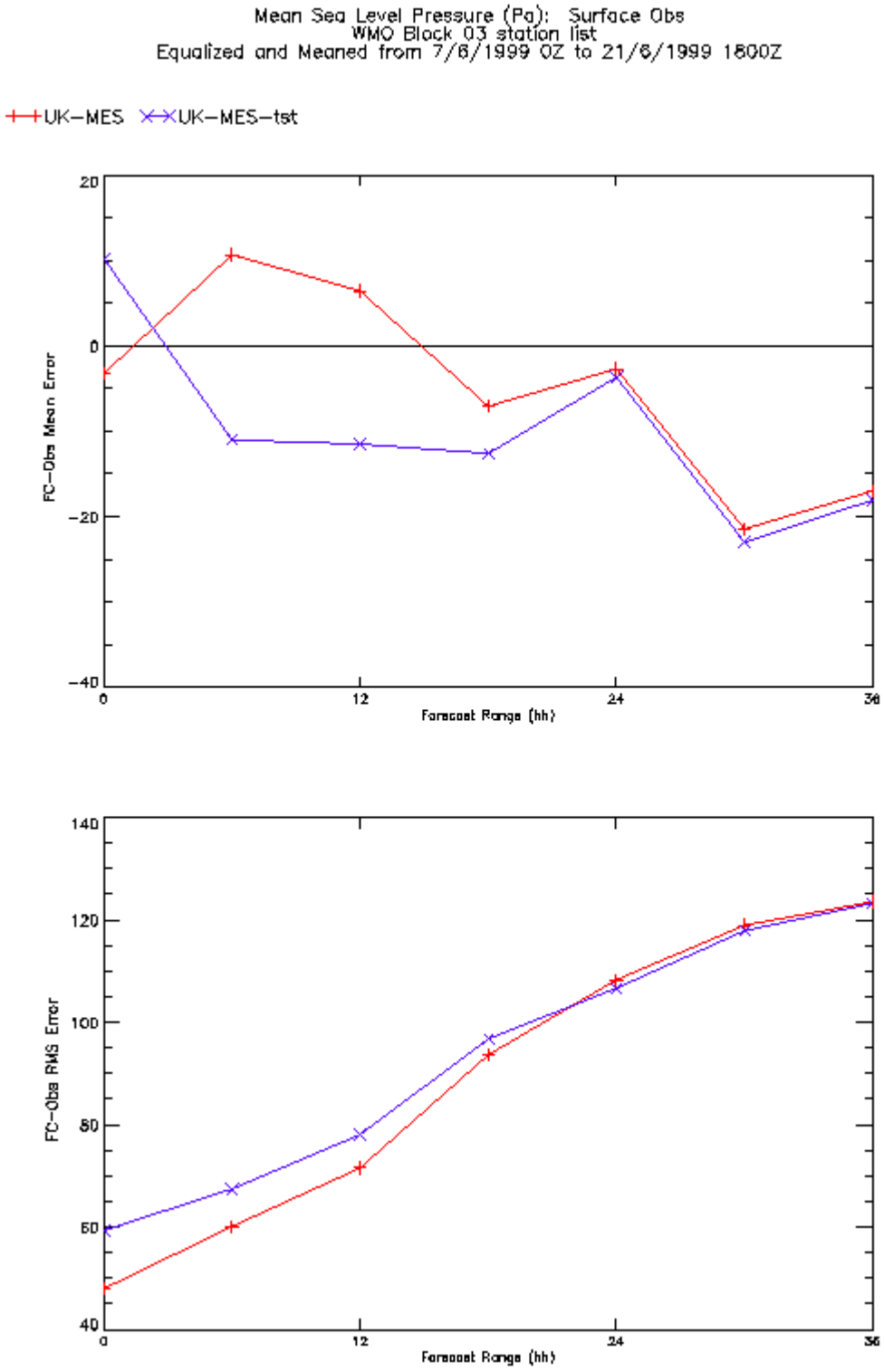

Figure 1.14: Verification of PMSL (Pa) for real-time phase of 3DVar implementation trial. Results averaged over all UK stations. Control in red, 3DVar in blue. Top frame shows mean error, lower frame shows rms error.

2. The June 2000 Upgrade to 3DVar

2.1 Specification

The components of the package were:

- New *forecast error covariance statistics*. These were based on summer and winter forecast data, instead of just winter data as in the original implementation. They use ''Rotated Kinetic Energy Vertical Modes' instead of the previous unrotated streamfunction and velocity potential vertical modes. This resulted in sharper vertical wind correlations in better agreement with observation. An error in the horizontal spectrum present in the first 3DVar version was corrected. This had given unrealistically short horizontal scales, especially in the upper-air. Horizontal scales for wind and temperature were increased still further, based on analysis of forecast difference statistics. The streamfunction and velocity potential variances were scaled to give a less divergent wind increment, which should be retained better by the model.
- Improved observation processing of surface pressure, as already implemented in the global model.
- Restored use of hourly SYNOPs, taking advantage of the availability of time interpolated model background.
- New stopping criterion in 3DVar descent algorithm to prevent excessive iteration count in some cases.
- Minor changes to the latent heat nudging assimilation of precipitation data: removal of the false contribution from ice advection to the large scale latent heating rate, a slightly shorter filter scale, and removal of divergence damping.
- Wind profiler data now extracted and processed for monitoring, with a view to admitting them to the assimilation at a later date.

2.2 The parallel trial

The 'real-time' parallel trial ran for a total of 2 weeks between 7th-26th April 2000.

2.3 Case studies

Sixteen cases were chosen, covering the required range of synoptic types. Each was run with 12 hours of assimilation followed by a forecast to t+36.

00Z 25/01/2000	radiation fog
00Z 13/11/1999	stratus
12Z 19/01/2000	stratocumulus
00Z 25/12/1999	organised convection – winter
12Z 16/08/1999	organised convection – summer
12Z 05/06/1999	convection - diurnal, local heating, land-based summer
00Z 05/11/1999	air mass (active)front
18Z 20/12/1999	mixed snow and rain
06Z 11/07/1999	clear summer day
12Z 26/01/2000	clear winter night
00Z 03/12/1999	land gales
06Z 29/06/1999	summer rain (trailing cold front affecting Wimbledon)
06Z 05/09/1999	hot ‘late summer’ day
12Z 18/12/1999	snow
00Z 16/03/2000	anticyclone
18Z 03/04/2000	heavy snow and rain

2.4 Subjective assessment

2.4.1 Case Studies

All 16 case studies were assessed for 3-hourly frames for each of the following variables: precipitation, cloud (high, medium and low), PMSL, 1.5m temperature, probabilities of fog, stratus, and snow (when relevant), and convective cloud fraction, all out to the maximum range of T+36. Unless stated otherwise the comparisons revealed either insignificant differences, or the differences were such that there was no clear preferred run.

Stratocumulus (19/1/00)
No overall preference for either run.

Organised summer convection (16/8/99)
The trial was slightly preferred for precipitation, but only at T+0, for having fewer showers over the Midlands.

Mixed snow and rain (20/12/99)
On the T+18 frame of the trial there was a reduction in the stratus probability over parts of East Anglia, and the control run’s higher probability was preferred.

Clear summer day (11/7/99)
No significant differences.

Clear winter night (26/1/00)
On the T+3 and T+6 frames the trial run had lower fog probabilities over the Midlands. The control’s higher probability was preferred since an area of fog persisted over the South Midlands throughout the evening. See Figure 2.1. On later frames there was no preference because the control then had the fog in the wrong place.

Radiation fog (25/1/00)
The trial was slightly preferred overall for 1.5m temperature, for having slightly higher temperatures over Southern England at T+15 and T+33,36 and also for having slightly lower temperatures at T+27 and T+30. The control run was preferred for Stratus probability but only on one frame, T+15, on which the trial appeared to prematurely lose the probability of Stratus over part of the Midlands.

Stratus case (13/11/99)
No significant differences throughout.

Organised convection (winter – 25/12/99)
There were differences in PMSL (up to 4hPa in association with an eastward-moving trough), especially at T+0, T+6 and T+15, but there was no preference for either run because the surface plotted U.K. charts showed the actual pattern to be about half way between the two forecast outputs.

Diurnal convection (5/6/99)
No preference.

Active front (25/11/99)
No preference.

Land gales (3/12/99)
Once again there were some significant PMSL differences, this time in a secondary small low developing off the Norwegian coast from T+30, but the actual pressure pattern was again somewhere between the two forecasts so there was no preference for either run.

Weak front (29/6/99 – the "Wimbledon case")
06Z run:
The trial was slightly preferred for having a slower clearance of the rain from SE England on the T+15 and T+18 frames. See Figure 2.2.
00Z run:
No preference.

Late summer day (5/9/99)
The trial was preferred on one frame, T+3, for retaining more fog over the north-eastern part of E Anglia.

Snow (18/12/99)
No preference.

Anticyclone (16/3/99)
No preference.

Mixed snow and rain (3/4/00)
Some differences on individual frames of precipitation and cloud, but no overall preference.

Summary of case studies

The case studies showed a broadly neutral signal. For the majority of forecast frames any differences between the control and upgrade runs were very small and for the most part any small forecast improvements were balanced by small detriments. On balance, however, there seemed to be a very slight overall improvement in forecasting of precipitation and a very slight detriment to forecasts of low cloud and fog.

2.4.2 Subjective assessment of parallel trial

At least one run was assessed from each of the following dates in April: 7th, 10th, 11th, 12th, 13th-21st inclusive, 25th, and 26th. The early dates (to the 14th) were assessed for all the variables as in the case studies. After this, due to time constraints, the forecast runs were only assessed for differences in precipitation.

7th April
The trial run was preferred at T+15 for having less rain over the Midlands, otherwise no preference.

10th April
No significant differences.

11th April
The control run was preferred on one frame, T+21, for having more rain over N Ireland and parts of S Wales, otherwise no preference. See this example in Figure 2.3.

12th April
The trial correctly had more low cloud at T+33 over parts of the Midlands, E Anglia and S England, which had the knock-on effect of correctly ~2-degree lower temperatures in these areas in the trial on this same forecast frame.

13th,14th,15th April
No preference.

16th April
At T+12 the trial was slightly preferred for a better orientation of a rain-band across England.

17th April
AT T+18 the trial was slightly preferred for having less precipitation over part of Ireland and E Scotland.

18th April
No preference.

19th April
At T+3 the trial was slightly better over E Scotland and at T+12 for a slightly faster spread of precipitation eastwards over Scotland.

20th April
No significant differences.

21st April
No significant differences.

25th April
No preference.

26th April
The control was preferred at T+0 because the trial run had less precipitation over E Anglia and more over the Midlands, which was not correct. There was no preference however for the subsequent frames.

Summary of real-time trial
The number of frames where there were significant differences, and a judgement could be made as to which run made the better or preferred forecast, were very small. It could be argued that overall there was a very slight preference for the trial runs.

2.5. Objective Verification

2.5.1 Case Studies

The variables making up the 5-component UK index were verified and the results presented in Table 2.1 are averaged over all cases and forecast times from t+6 to t+24, as used in the index. The station list used is 'all UK stations', which provides better sampling than the standard UK index list of only ~40 stations.

Variable	Control		Trial		Change in Skill/ETS
	rms	ETS	rms	ETS	
10-metre wind	2.88	0.750	2.90	0.747	-0.003
screen temperature	1.59	0.650	1.59	0.651	+0.001
precipitation		0.379		0.375	-0.004
cloud amount		0.227		0.224 }	
				}	+0.001
cloud base		0.139		0.145 }	
visibility		0.129		0.124	-0.005
average 'UK index'		0.418		0.416	
		%age change in UK index for trial			-0.5%

Table 2.1 - Summary of UK index components for the case studies

Skill changes for *10-metre wind and screen temperature* were estimated on the basis of mean skill in the control being 0.75 for wind and 0.65 for temperature.

For *precipitation*, the verification is against 6-hourly accumulations at SYNOP stations, as verification against Nimrod data was not yet convenient for case studies. The scores were derived from a single contingency table per threshold, collating events at all forecast times. This is a more stable way to get an average figure than averaging the ETS values from separate tables for each forecast time at each threshold. (The reason for this is that 6-hourly accumulation observations are available only at 00Z and 12Z, so that if the distribution of forecast data times for the chosen cases is not even, some forecast times have few verification data and may give unreasonable contributions to an average which gives equal weight to each forecast time.)

The very slight detriment in the trial ETS is made up of a benefit for the 4mm threshold and detriments for the 1mm and 0.2mm thresholds, but it must be noted that only around 10 of the 16 cases had significant rainfall, and the surface network is sparse for deriving a reliable signal from such a sample. Verification against radar data, as performed for the real-time trial is likely to be a better guide to long term performance. We note that subjective assessment of the cases found a marginal improvement in the trial precipitation.

For *cloud*, the change in ETS derives from the mean of scores for cloud amount and cloud base. Verification of the mean cloud cover error showed the trial to have a slightly reduced negative bias in the first 12 hours.

For *visibility*, although the ETS figures for UK stations imply a slight detriment from the trial, there are few events at the lower two thresholds. The results for the whole model area were 0.114 for the control and 0.115 for the trial, so a neutral signal is expected.

Summary

Table 2.1 shows an estimated UK Index impact of just –0.5%. For both the list of 'all stations in the model area' and the UK Index list, there was no overall impact to 3 significant figures, so the overall trial impact is best described as neutral.

2.5.2 Parallel trial

The variables making up the 5-component UK index were verified for all available runs over the period 7th-22nd April 2000. Results are averaged over forecast times from t+6 to t+24 as used in the index, using a station list of 'all UK stations'. The results are summarised in Table 2.2.

Variable	Control		Trial		Change in Skill/ETS
	skill	ETS	skill	ETS	
10-metre wind	0.785		0.795		+0.010
screen temperature	0.683		0.689		+0.006
precipitation		0.246		0.259	+0.013
cloud amount		0.168		0.174 }	
				}	+0.005
cloud base		0.111		0.116 }	
visibility		0.060		0.066	+0.006
average UK index		0.383		0.391	
		%age change in UK index for trial			+2%

Table 2.2 - Summary of UK index components for the parallel trial

For *screen temperature*, the slight improvement in skill score is evident in reduced rms errors in the later stages of the forecast, even beyond t+24 (Figure 2.4).

For *precipitation*, verification was derived from comparison of the model accumulations with those from Nimrod analyses. ETS results are shown in Figure 2.5. They show a systematic small improvement in the trial results at each threshold and forecast time. It is noteworthy that the improvement is as large at t+24 as earlier in the forecast, and is slightly greater for heavier thresholds. The most likely explanation for the improved precipitation signal is an improvement in the wind analysis coming from the covariance changes.

The Frequency Bias for precipitation events at different thresholds was also examined. For the lightest threshold at t+6, the trial brings a 2.6% increase (detriment) relative to the operational bias of 1.11. For the heaviest threshold at t+6, however, the trial brings a 6% increase (benefit) relative to the operational figure of 0.81. Changes in bias later in the forecast were smaller.

There were not many UK observations of fog in the real-time trial, but the objective *visibility* results for the whole model area also showed a very slight advantage to the trial.

Summary

The balance of objective verification for the parallel trial favoured the trial, implying a 2% improvement in the UK index. All variables showed a slight improvement.

2.6 Summary of June 2000 3DVar Upgrade

A trial of an upgraded version of 3DVar in the UK Mesoscale model was completed, involving 2 weeks of 'real-time' parallel running alongside the operational system in April 2000 and 16 case studies covering a variety of synoptic types.

Of the numerous package ingredients, the most significant was an improved representation of forecast error covariances, giving longer horizontal scales and shorter vertical scales for the analysis increments than those initially implemented in October 1999. This development also

gave potential for further improvement from distinguishing the scales appropriate for surface and upper-air analysis.

A subjective assessment of the *'real-time' phase* found only few and slight differences but with an overall very slight preference for the trial. Overall objective verification of the real-time trial suggested a 2% increase in the estimate of the 5-component UK index for the trial, with a systematic slight improvement in precipitation forecasts at each threshold and forecast time up to t+24. A subjective assessment of the *case studies* returned a broadly neutral signal but on balance, however, there seemed to be a very slight overall improvement in forecasting of precipitation and a very slight detriment to forecasts of low cloud and fog. The objective verification of the case studies showed a trial 5-component UK index that was neutral with respect to the operational run.

Taking the trial as a whole, the overall signal from the June 2000 upgrade package is probably best described as a slight improvement for precipitation and neutral for other variables.

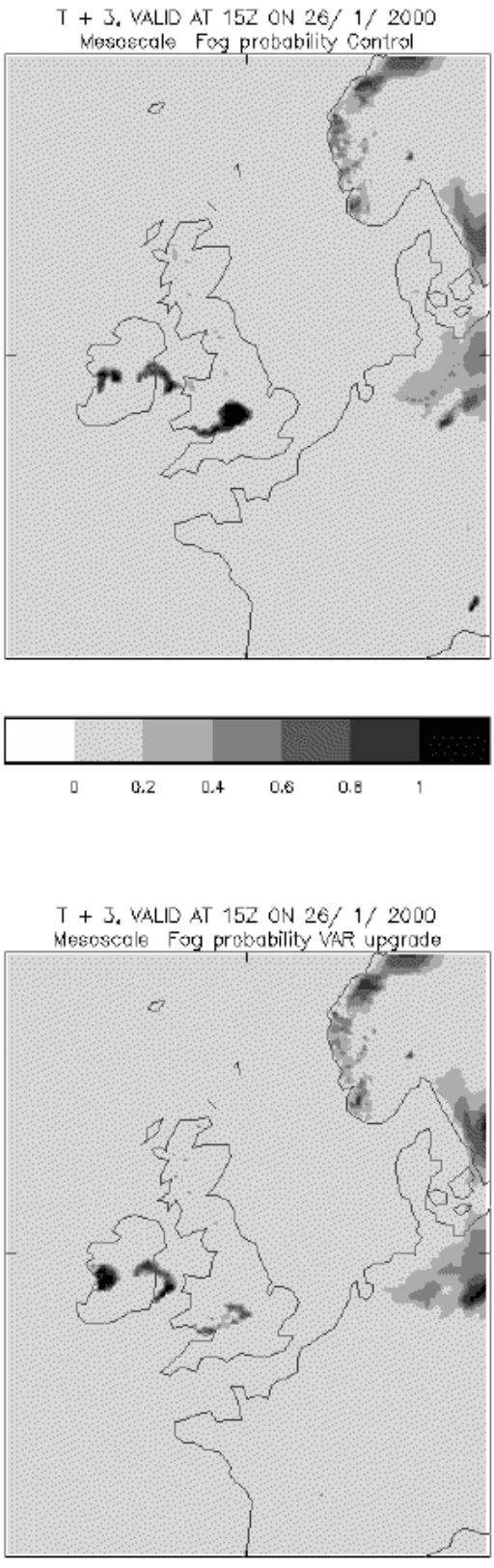

Figure 2.1: Case study trial of June 2000 3DVar Upgrade. An example of the Control run (top panel) being preferred to the Trial run (bottom panel), in this case for keeping a higher fog probability at t+3 from DT 12 UTC 26/1/00, in an area where fog was occurring.

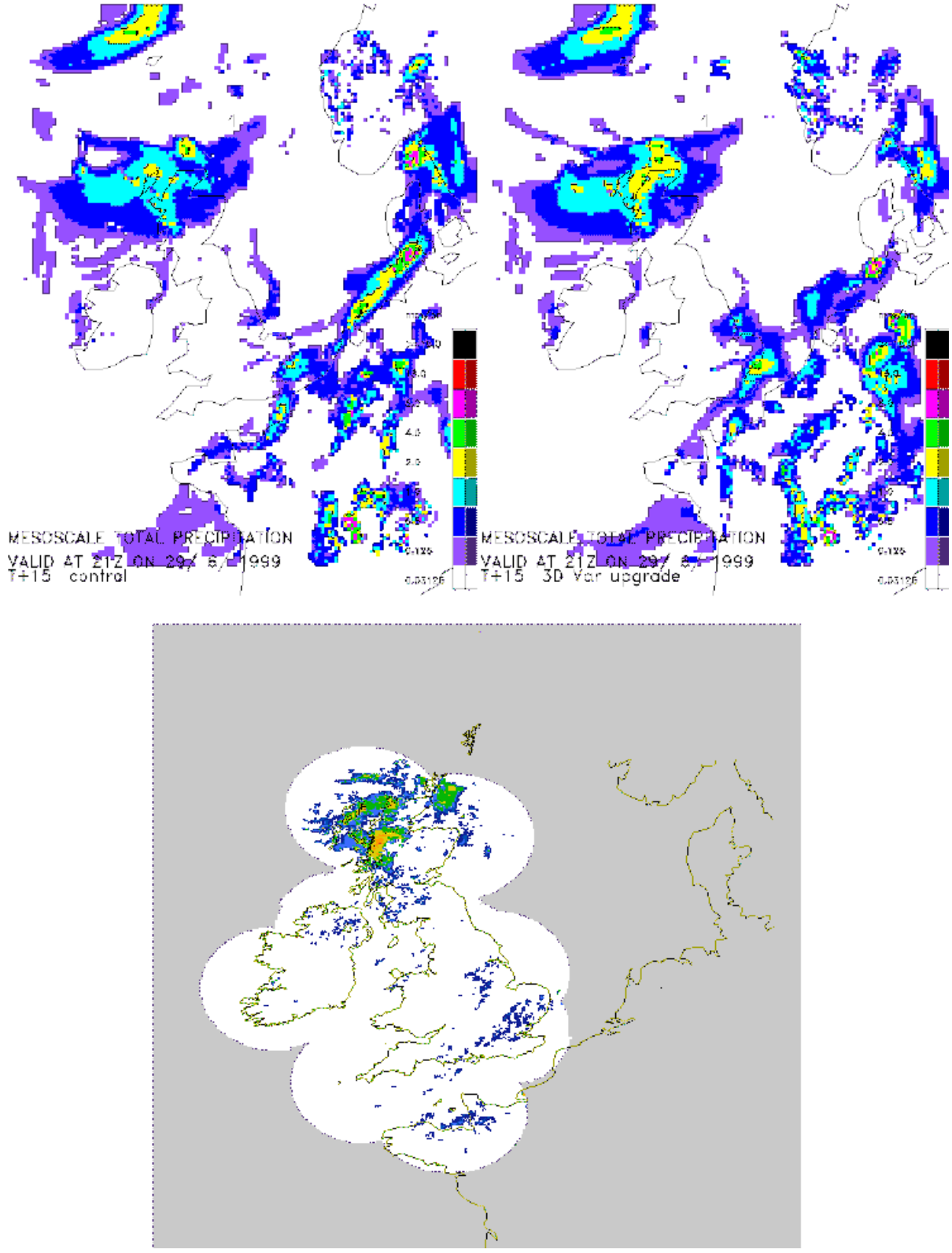

Figure 2.2: Case study trial of June 2000 3DVar Upgrade. T+15 rain rate forecasts for the 'Wimbledon' case, DT 06 UTC 29/6/99: Control (top left) and Trial (top right), with verifying radar in bottom frame. The Trial run's slower clearance of rain from SE England is preferred.

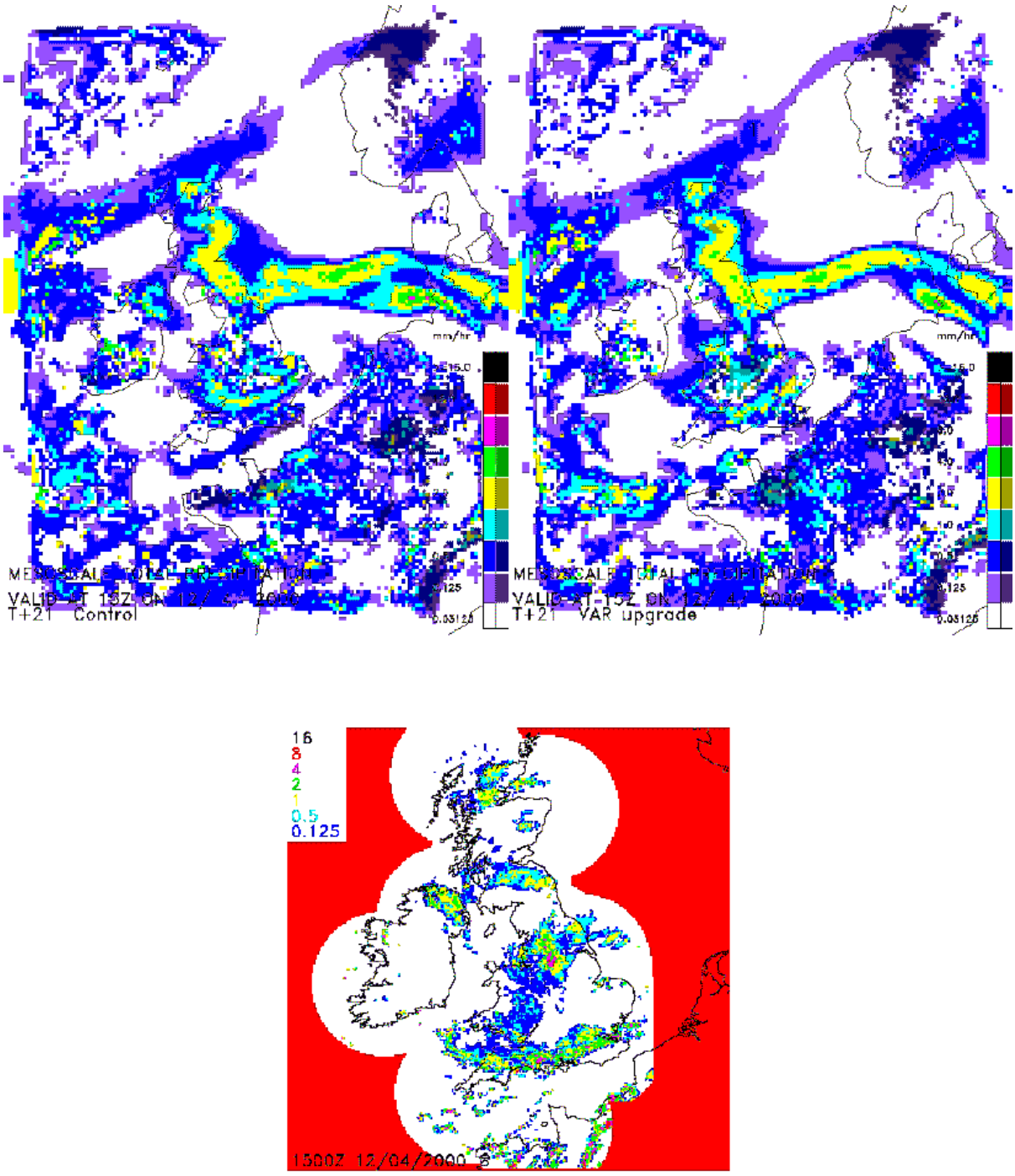

Figure 2.3: An example from the real-time trial of the June 2000 3DVar upgrade, t+21 rain rate forecasts for VT 15 UTC 12/04/2000. The Control run (top left) is preferred to the Trial (top right), with more rain correctly over part of N. Ireland and S. Wales, as revealed by verifying radar composite (lower panel).

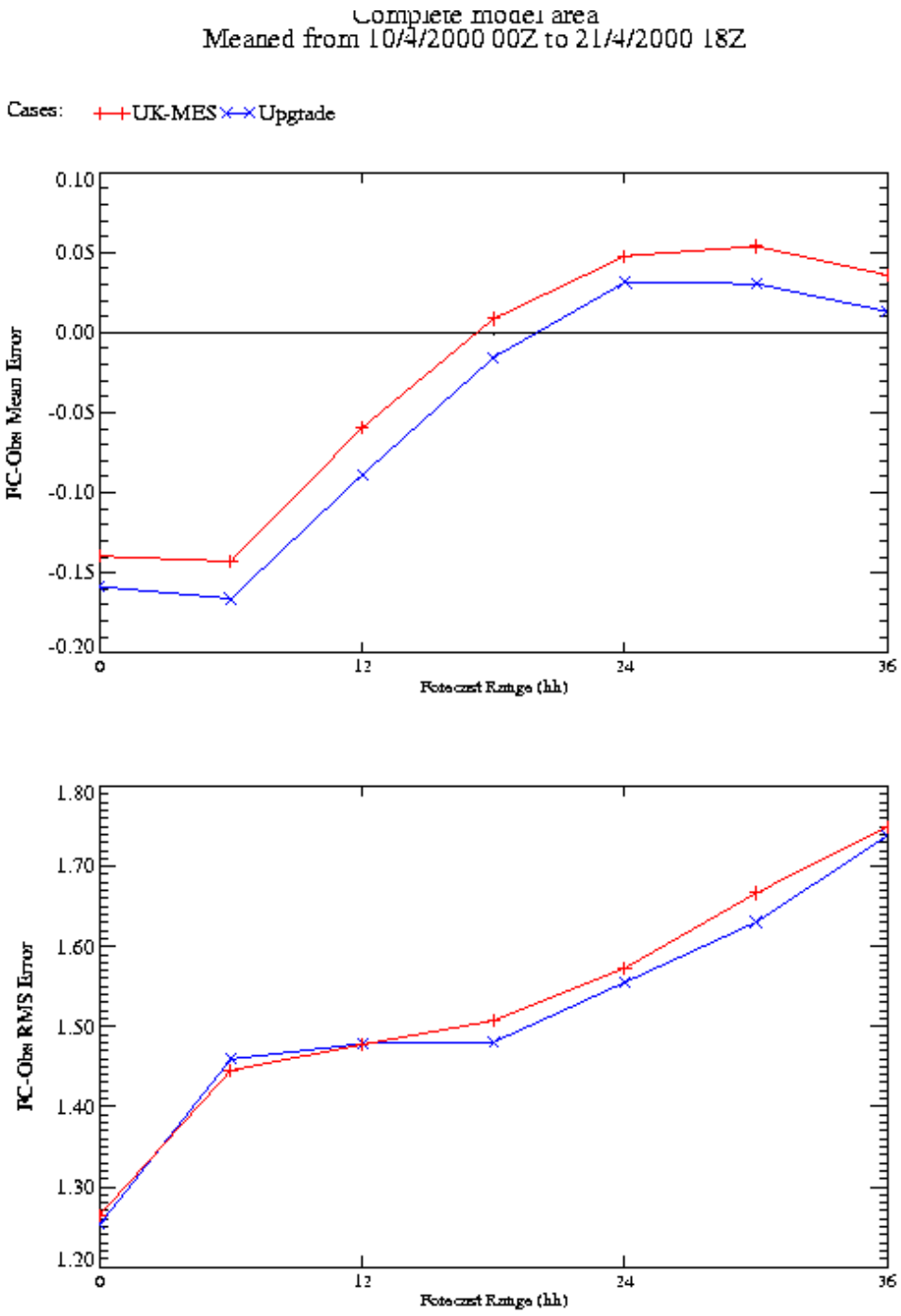

Figure 2.4: Screen temperature errors from the real-time trial of June 2000 3DVar upgrade. Top panel shows mean errors, bottom panel shows rms errors. Red lines for Control and blue lines for Trial.

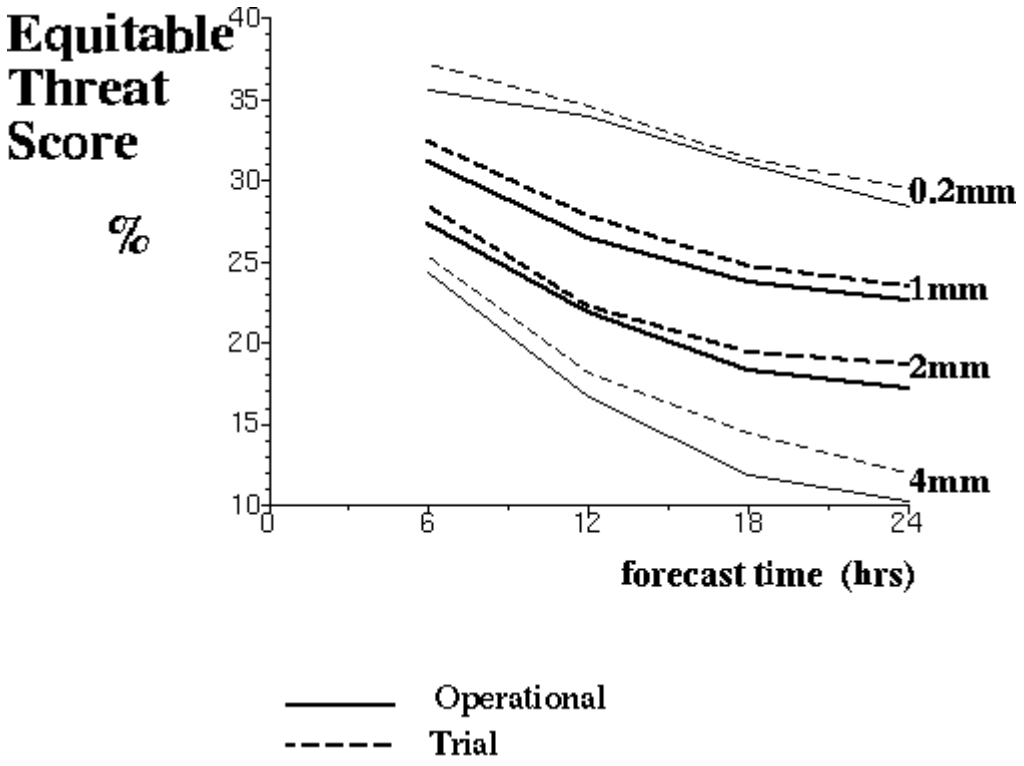

Figure 2.5: Verification of 6-hourly accumulations v Nimrod for the real-time trial of the June 2000 3DVar upgrade (35 forecasts from 7th-22nd April 2000), at 4 different thresholds.

3. The February 2001 Upgrade to 3DVar

3.1 Specification

The components of the package were:

- Introduce a *soil temperature increment* equal to the atmospheric temperature increment at the lowest model level. The increment is added to the surface temperature, T_* and the top deep soil temperature level. An increment like this used to be part of the old AC scheme, though its implementation here is modified slightly by the detail of the MOSES II 'tile scheme' edition of the land surface scheme.

- Introduce assimilation of selected UK and European *wind profiler data*. An impact trial of wind profilers in the global model had already proved satisfactory. Continued monthly monitoring and comparison of (observation-model) statistics with nearby sondes helped to define a suitable station list.

  All profiler data is thinned in time to one report per hour. The Aberystwyth PILOT is switched-off and the BUFR wind profiler (03501) used above 700hPa. The 3 UK boundary layer profilers at Aber (03500), Camborne (03807) and Dunkeswell (03840) are used below 500 hPa. We also began assimilation of the *Swedish weather radar VAD profiler* at Leksand (02430) below 500 hPa, the European profilers at Cabauw (06348), Payerne (06610) and La Ferte Vidame (07112, above 700 hPa), but not the other French profiler at Clermont Ferrand (07453) due to significant biases. There are three observation error profiles specified. Most profilers are given the lowest errors, but Payerne and La Ferte Vidame are assigned slightly higher errors. The initial probability of gross error, PGE = 0.02 generally, but is higher (0.1) for the two profilers with the higher observation errors. (Since the profiler data were introduced, monitoring has continued, in order to refine the station list. There is now also a profiler at Wattisham in eastern England.)

- Introduce extra high density *Meteosat wind data*. Previously we used low-resolution products derived from the IR, VIS and WV channels, and in the mesoscale model only data from METEOSAT-7. These were used unthinned and quality control rejected typically 20% of the data, leaving usually around 20 observations per run. EUMETSAT make available additional winds with higher spatial resolution, higher temporal resolution (90 mins) and quality indicators from the EUMETSAT processing. Coverage after quality control is improved to 100+ data per run, including more in the Biscay area. Most of the extra data are in the 100-400hPa band. The rejection criteria set are:

  1. all winds rejected above 100hPa
  2. WV rejected below 400hPa
  3. VIS rejected above 700hPa
  4. VIS rejected over land
  5. IR rejected below 700hPa over land

  A week long real-time trial of this change alone showed a neutral impact.

- *Correction of a wind rotation error in the OPS and Var steps*. In the quality control step, observed winds relative to true latitude-longitude were compared with model winds still relative to a rotated latitude-longitude grid. In the Var step, model background winds were erroneously rotated a second time to the model grid. In a ten day real-time trial of the Var step correction alone, the impact on the UK index was neutral.

- Correction of a bug which led to certain *quality control parameters* having values appropriate to the global model, instead of the intended mesoscale values.

- Correction of a one grid length displacement error in interpolation of the Nimrod cloud analysis to the model grid.

- Introduction of the *merge back of associated model data for total cloud cover to the MetDB*, for monitoring purposes only.

3.2 The parallel trial

The 'real-time' parallel trial ran from 22nd December 2000 to 11th January 2001.

3.3 Case studies

Twelve cases were chosen, covering the required range of synoptic types. Each was run with 12 hours of assimilation followed by a forecast to t+36.

00Z 25/01/2000	radiation fog
06Z 05/06/2000	stratus
12Z 19/01/2000	stratocumulus
00Z 25/12/1999	organised convection – winter
18Z 03/06/2000	organised convection – summer
12Z 18/05/2000	convection - diurnal, local heating, land-based summer
06Z 27/09/2000	air mass (active)front
18Z 20/12/1999	mixed snow and rain
06Z 21/07/2000	clear summer day
12Z 26/01/2000	clear winter night
00Z 03/12/1999	land gales
06Z 29/06/1999	summer rain (trailing cold front affecting Wimbledon)
00Z 03/12/1999	hot 'late summer' day

3.4 Subjective assessment

3.4.1 Case study trial

All twelve case studies were assessed against current operational mesoscale output, with subjective comparison against plotted surface charts, radar and, where necessary, satellite imagery. All runs were compared on 3-hourly frames for the usual fields:

Out of the 12 runs, 8 were found to have no significant differences in any of the fields compared. The remaining 4 cases and their differences are described below:

00Z 25th January 00 - patchy radiation fog
On the T+21 frame the trial run had a very slightly better distribution of high fog probability generally.

12Z 19th January 00 - stratocumulus
On only the T+6 frame the trial run had a little more low cloud over central southern England which was preferred. On only the T+21 frame the trial run had a lower probability of stratus over the south and southwest of England, this was also preferred. The 50% cover threshold of low cloud is shown in Figure 3.1 below.

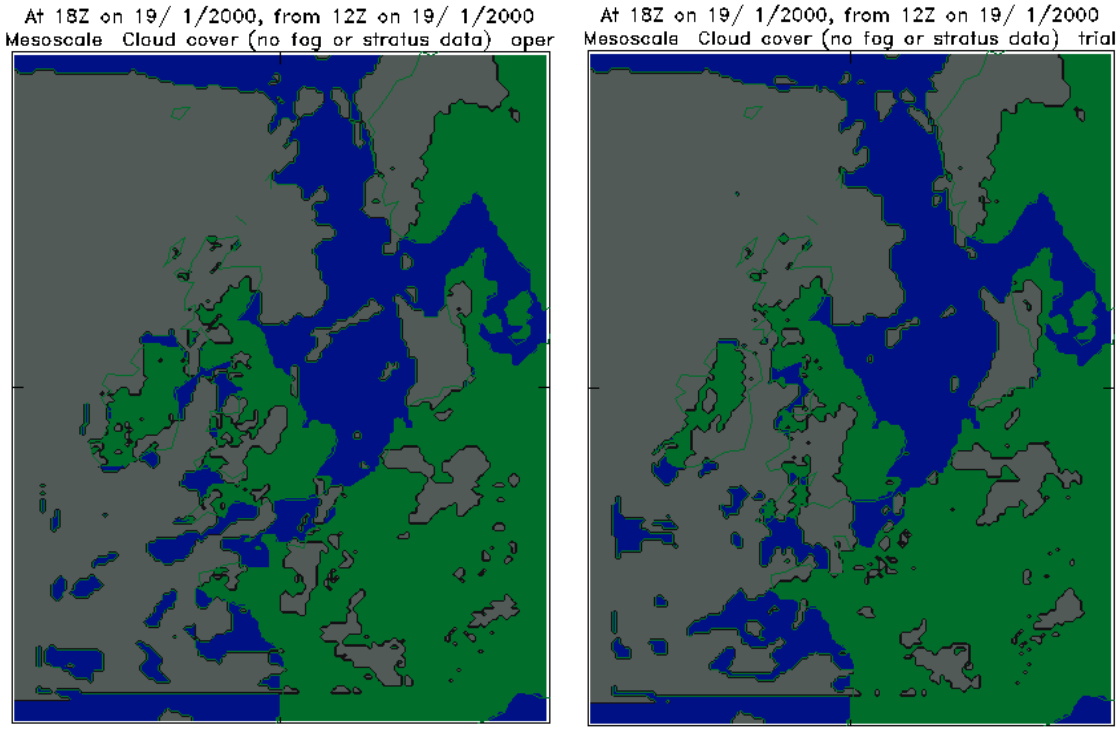

Figure 3.1: February 2001 3DVar upgrade trial - t+6 cloud cover from DT 12 UTC 19/1/00. Operational forecast (left frame) and trial forecast (right frame)

18Z 3rd June 00 - organised convection
At T+0 the precipitation fields had some differences, with the trial being preferred over eastern Ireland and the operational run preferred over Somerset, but with no preference overall.

18Z 20th December 99 - mixed snow and rain
For precipitation there were some minor differences on a couple of frames but none such that there was a preference for either run. For PMSL, however, the trial run was preferred at T+0 for having the 1026 and 1028hPa isobars better placed over eastern England, as shown in Figure 3.2 below.

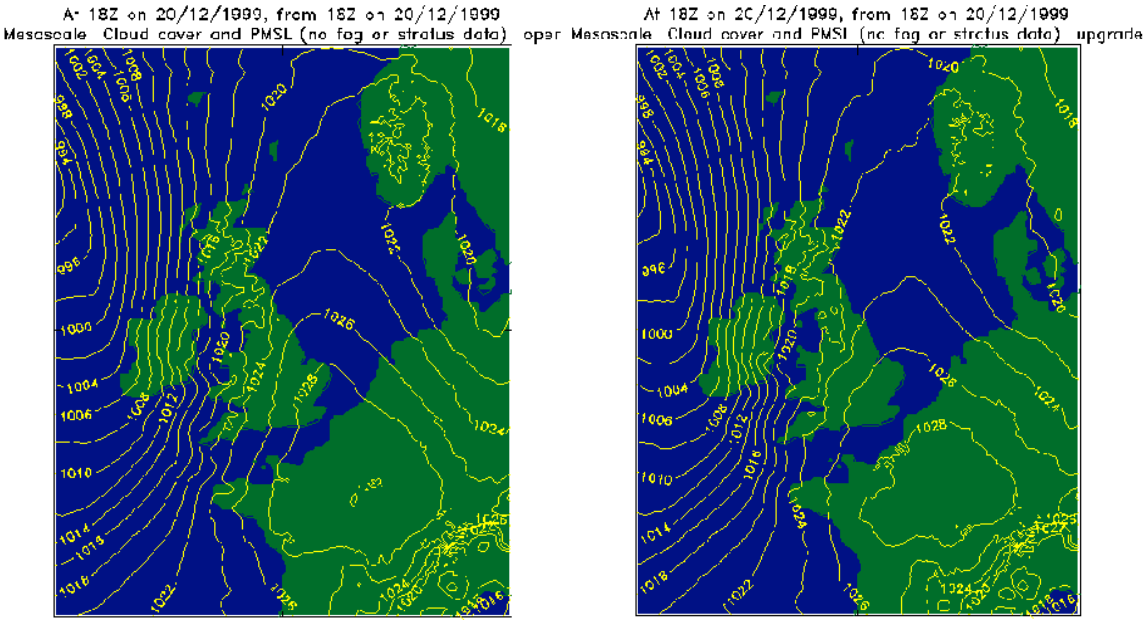

Figure 3.2: February 2001 3DVar upgrade trial - t+0 PMSL from DT 12 UTC 20/12/99. Operational forecast (left frame) and trial forecast (right frame)

3.4.2 Subjective assessment of the parallel runs

Sixteen pairs of runs (current operational and trial) were assessed from the period 4th to 21st January 2001, for the same fields as in the case studies. For the majority of runs there were no significant differences for most or all the fields examined. For the 5 runs where there were significant enough differences in one or more fields for a preference to be stated, there were 2 where the operational run was favoured and 3 where the trial output was preferred, but in all 5 the preference can only be described as slight. Descriptions of the differences where there was a preference:

00Z 4th January:
for precipitation the operational run was slightly preferred at T+6 for having less spurious rain over part of southern Scotland and the Irish Sea. This is shown in Figure 3.3 below.

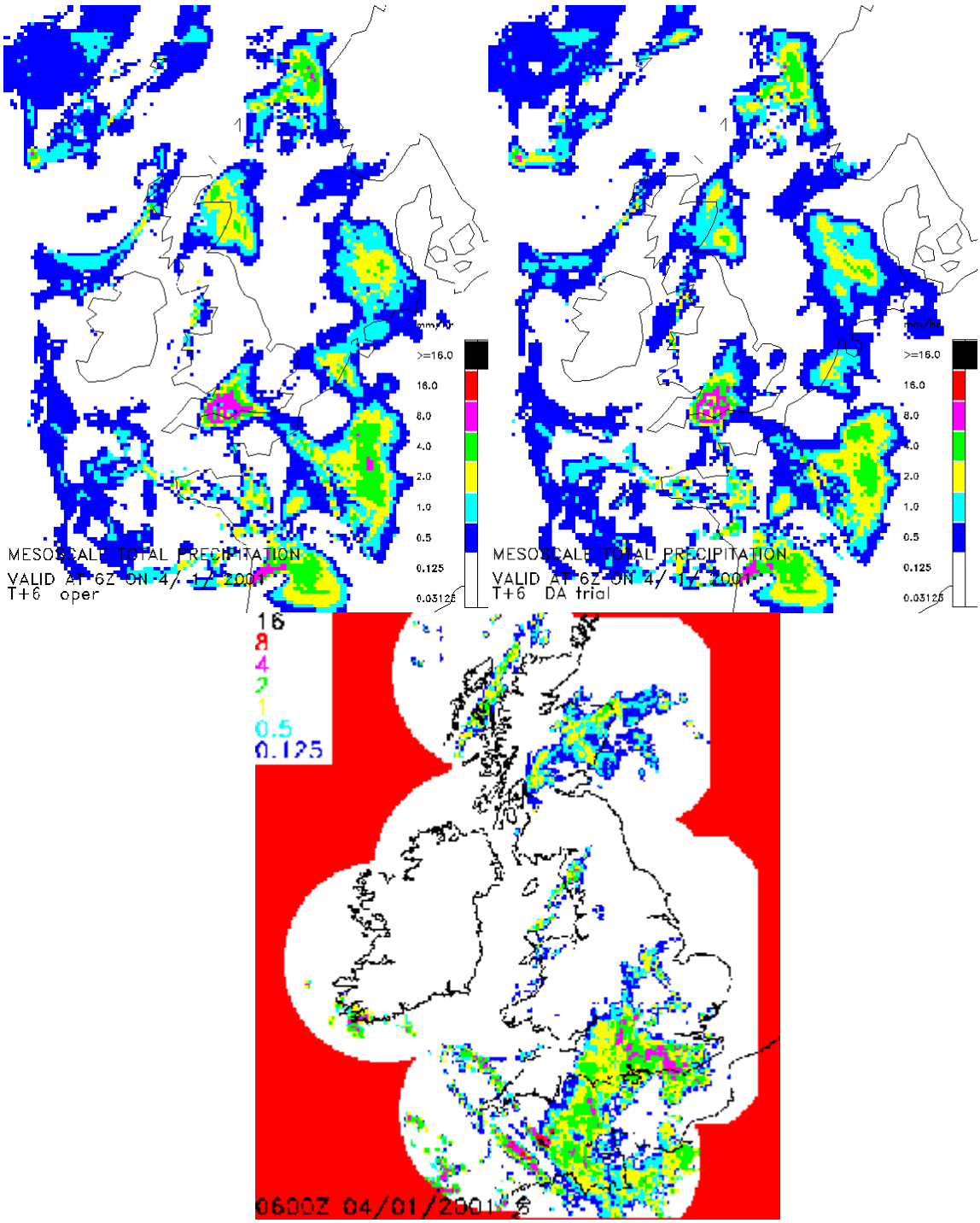

Figure 3.3: February 2001 3DVar upgrade trial - t+6 rain rate from DT 00 UTC 4/1/01. Operational forecast (top left frame) and trial forecast (top right frame). Verifying radar composite (lower frame).

00Z 5th January:
on the T+21 frame the trial run was preferred for having less spurious rain over the Midlands and for having lower fog and stratus probabilities over the Midlands and northern England on the T+12, 15, 18 frames.

12Z 6th January:
trial preferred for having somewhat lower stratus probabilities on later frames over SE England, post cold front.

00Z 9th January:
trial preferred for being slightly warmer than the operational run at T+0 and T+6 and for having a lower fog probability over SE England at T+6.

00Z 21st January:
operational run preferred for being slightly colder and with a higher (~0.2) fog probability over parts of Ireland and England from T+15 onwards.

Summary of subjective assessment
There were few occurrences of significant differences between current operational chart output and trial output, but overall there was a slight preference for the trial charts in both the case studies and the parallel runs.

3.5 Objective Verification

3.5.1 Case Studies

The variables making up the 5-component UK index were verified and the results presented in Table 3.1 are averaged over all cases and forecast times from t+6 to t+24, as used in the index. The station list used is 'all UK stations', which provides better sampling than the standard UK index list of only ~40 stations.

Variable	**Control**		**Trial**		**Change in Skill/ETS**
	rms	**ETS**	**rms**	**ETS**	
10-metre wind	2.89	0.750	2.90	0.750	0.000
screen temperature	1.51	0.650	1.48	0.662	+0.012
precipitation		0.363		0.353	-0.010
cloud amount		0.245		0.245	0.000
visibility		0.109		0.105	-0.004
average 'UK index'		0.4234		0.4230	
		%age change in UK index for trial			-0.1%

Table 3.1 - Summary of UK index components for the case studies

Skill changes for *10-metre wind and screen temperature* were estimated on the basis of mean skill in the control being 0.75 for wind and 0.65 for temperature. For temperature, the reductions in rms error seen in the trial were at most 2.5% (Figure 3.4), compared with 4% in the real-time trial. It is likely that the kind of blocked winter pattern affecting a large part of the real-time trial

is more conducive to a signal from the soil temperature increment change. Although screen relative humidity is not part of the UK index, it also shows reduced errors in the trial (Figure 3.5).

For *precipitation*, the verification is against 6-hourly accumulations at SYNOP stations, as verification against Nimrod data is not yet convenient for case studies. The very slight detriment in the trial ETS is made up of a benefit for the 0.2mm threshold and detriments for the 1mm and 4mm thresholds, but it must be noted that only 6 of the 12 cases had significant rainfall, and the surface network is sparse for deriving a reliable signal from such a sample. We note that subjective assessment of the cases found a neutral signal for precipitation. Verification against radar data, as performed for the real-time trial, is likely to be a better guide to long term performance - the trial showed up slightly better overall by this measure.

For *visibility*, although the ETS figures for UK stations imply a slight detriment from the trial, the results for the whole model area (with about twice as many events) were 0.088 for both trial and control, so a neutral signal is expected. This is consistent with the real-time trial which had a larger number of fog events than the case studies.

Summary

Table 3.1 shows an estimated UK Index impact of just –0.1% i.e. neutral. There is a benefit in screen temperature and in screen relative humidity, agreeing with the real-time trial. There is a slight detriment in precipitation, but this is based on very few runs and did not show up subjectively, whereas the real-time trial with many more forecasts showed a slight advantage to the trial precipitation overall.

3.5.2 Parallel trial

The variables making up the 5-component UK index were verified for all available runs over the period 22nd December 2000 - 11th January 2001. Results are averaged over forecast times from t+6 to t+24 as used in the index, using a station list of 'all UK stations'. The results are summarised in Table 3.2.

Variable	**Control**		**Trial**		**Change in Skill/ETS**
	skill	**ETS**	**skill**	**ETS**	
10-metre wind	0.715		0.717		+0.002
screen temperature	0.624		0.632		+0.008
precipitation		0.287		0.290	+0.003
cloud amount		0.228		0.228	0.0
visibility		0.125		0.128	+0.006
average UK index		0.396		0.399	
		%age change in UK index for trial			+0.8%

Table 3.2 - Summary of UK index components for the parallel trial

For *screen temperature*, the improvement in skill score is evident in reduced rms errors out to t+24 (Figure 3.4). The rms error reduction is 4% at t+6 and 1% at t+24. There is an associated warming in the trial. Screen relative humidity errors are also lower in the trial (Figure 3.5)

For *precipitation*, verification was derived from comparison of the model accumulations with those from Nimrod analyses. ETS results are shown in Figure 3.6. They show an overall slight advantage for the trial with 7/16 scores better and only 4/16 scores worse for the various thresholds and forecast times.

Summary

Objective verification for the parallel trial favours the trial and implies ~1% improvement in the UK index. The main improvement is in screen temperature, as found in the case studies, with a slight overall benefit in precipitation.

3.6 Summary of February 2001 3DVar Upgrade

A trial of an upgraded version of 3DVar in the UK Mesoscale model was completed, involving 3 weeks of 'real-time' parallel running alongside the operational system and 12 case studies covering the required variety of synoptic types.

The package included: introduction of a soil temperature increment, assimilation of selected UK and European wind profiler data, assimilation of higher density Meteosat wind data and some minor bug fixes including the correct handling of wind components rotated from true north to model grid north.

Overall *objective verification of the real-time trial* suggested a 1% increase in the estimate of the 5-component UK index for the trial, with a systematic improvement in screen temperature forecasts up to t+24, the reduction in rms errors being 4% at t+6 and 1% at t+24. Screen relative humidity (not part of the index) was also improved.

The *objective verification of the case studies* showed a trial 5-component UK index neutral with respect to the control run, but with a 2% reduction in rms screen temperature errors.

In *subjective assessment*, there were few occurrences of significant differences between current operational chart output and trial output, but overall there was a slight preference for the trial charts in both the case studies and the parallel runs.

Taking the real-time and case study phases together, the overall signal from the trial package is best described as an improvement for screen temperature and neutral for other variables.

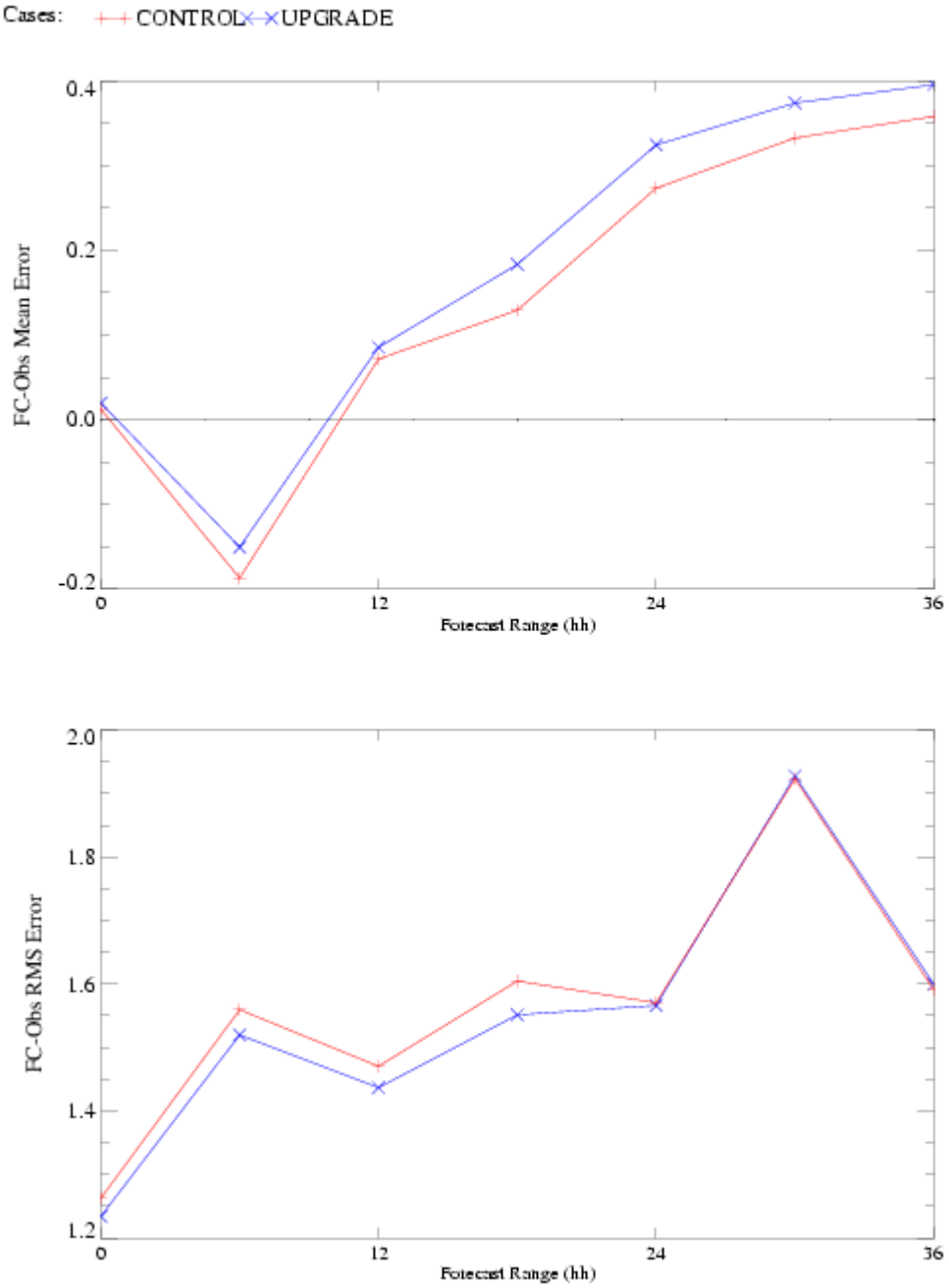

Figure 3.4: Screen temperature verification against 'all UK stations' from the case study trial of the February 2001 3DVar upgrade. Top panel for mean errors, bottom panel for rms errors. The trial, labelled 'upgrade' is in blue and the control is in red.

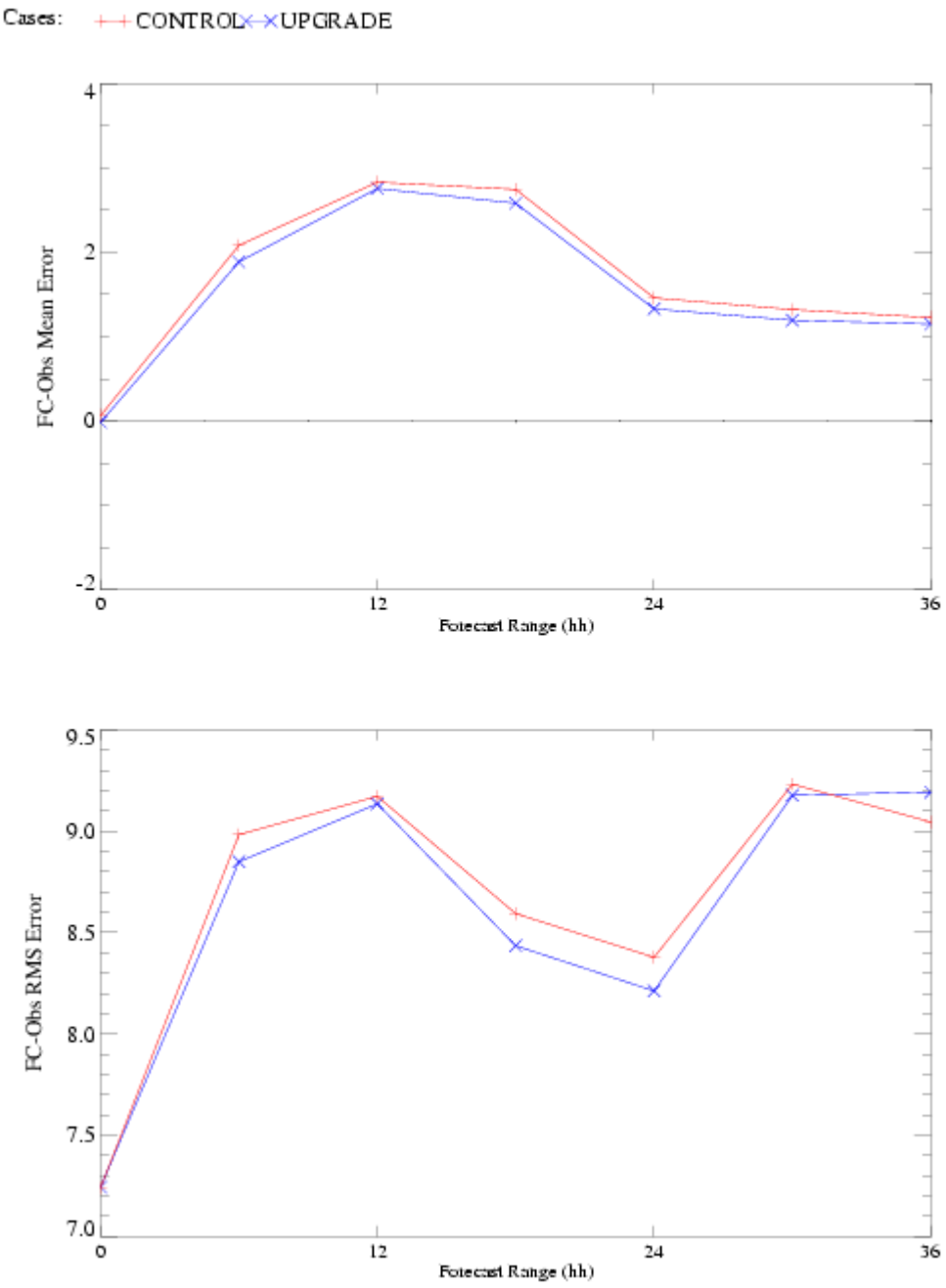

Figure 3.5: Screen relative humidity verification against 'all UK stations' from the case study trial of the February 2001 3DVar upgrade. Top panel for mean errors, bottom panel for rms errors. The trial, labelled 'upgrade' is in blue and the control is in red.

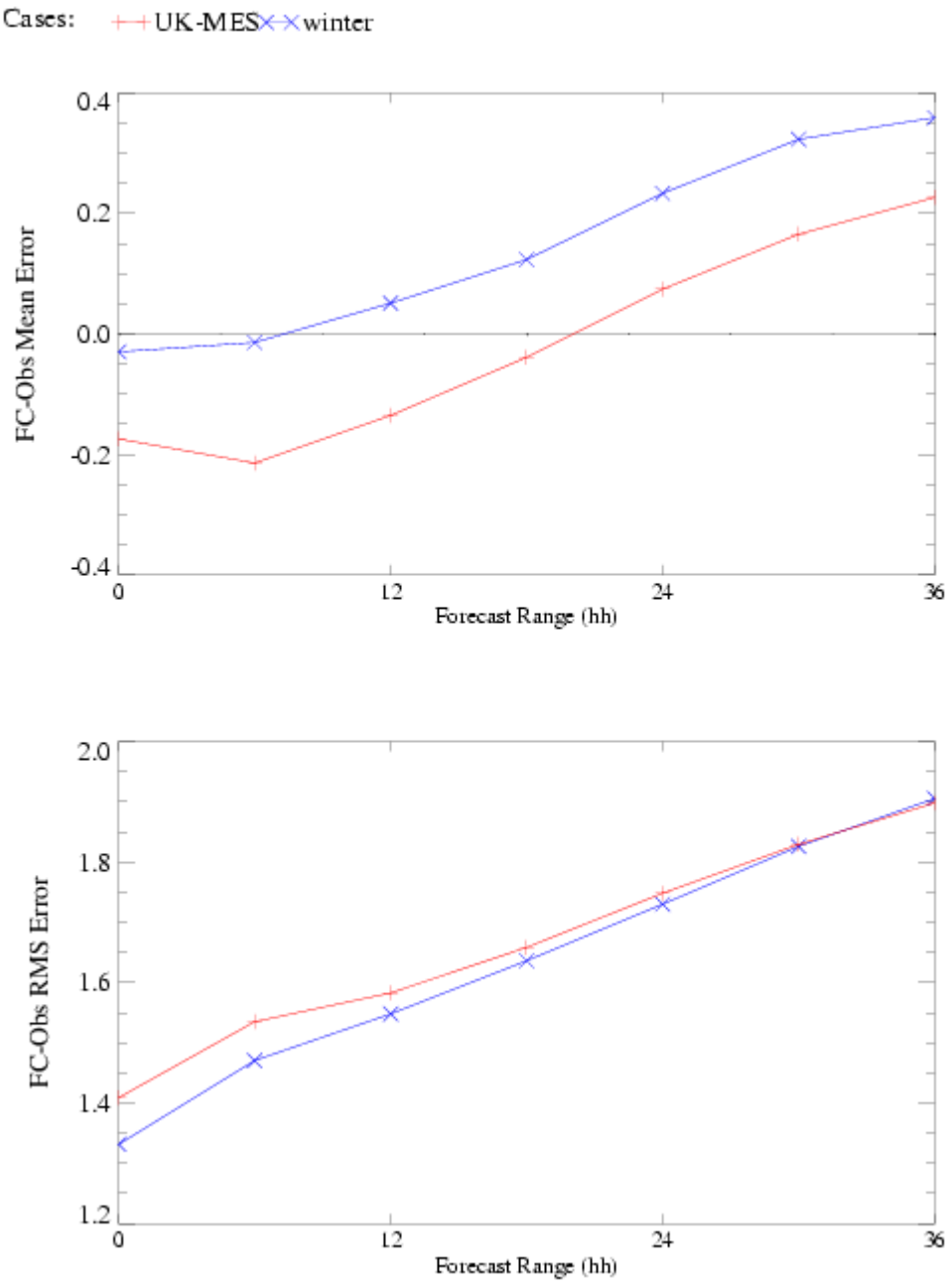

Figure 3.6: Screen temperature verification against 'all UK stations' from the real-time trial of the February 2001 3DVar upgrade. Top panel for mean errors, bottom panel for rms errors. The trial, labelled 'winter' is in blue and the control is in red.

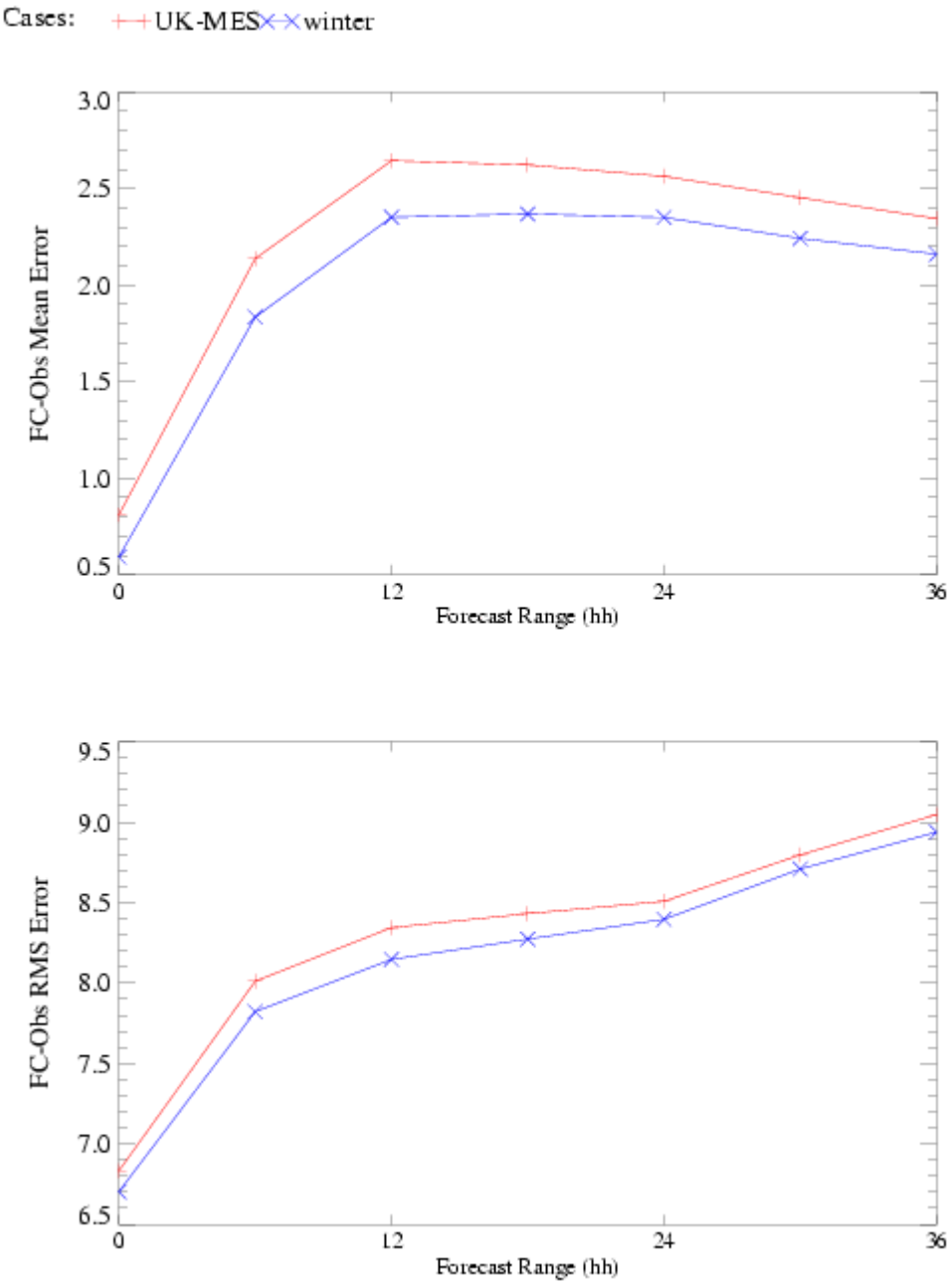

Figure 3.7: Screen relative humidity verification against 'all UK stations' from the real-time trial of the February 2001 3DVar upgrade. Top panel for mean errors, bottom panel for rms errors. The trial, labelled 'winter' is in blue and the control is in red.

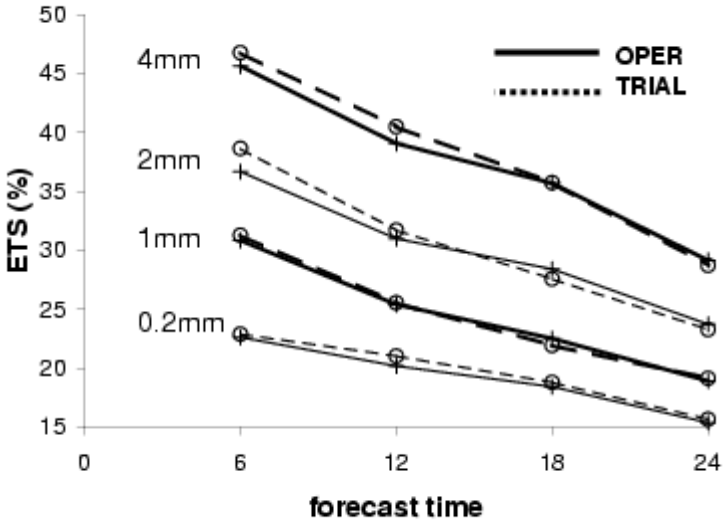

Figure 3.8: Verification of 6-hourly precipitation accumulations v Nimrod analyses for the real-time trial of the February 2001 3DVar upgrade (34 forecasts from 22nd December - 11th January). Equitable Threat Score (ETS) as a function of forecast time. For clarity, the ETS results for each of the 4 thresholds have been shifted by a different constant.

4. Developments in cloud assimilation

4.1 Introduction

During 1999, operational verification of total cloud cover showed increasing negative bias in both analysis and forecast (Figure 4.1). A debate arose over the source of the bias – had it come from changes in assimilation, or changes in model physics? To help answer this question, some observation impact experiments were run parallel to the operational suite for a 2 week period (17th-31st January 2000). In Experiment 1 (NoSondeRH), no radiosonde relative humidity (rh) data were assimilated, while in Experiment 2 (NoMOPSRH), no MOPS cloud data were assimilated. For a schematic of the MOPS system, and how it supplements the conventional data input to the mesoscale assimilation, see Figure 4.2.

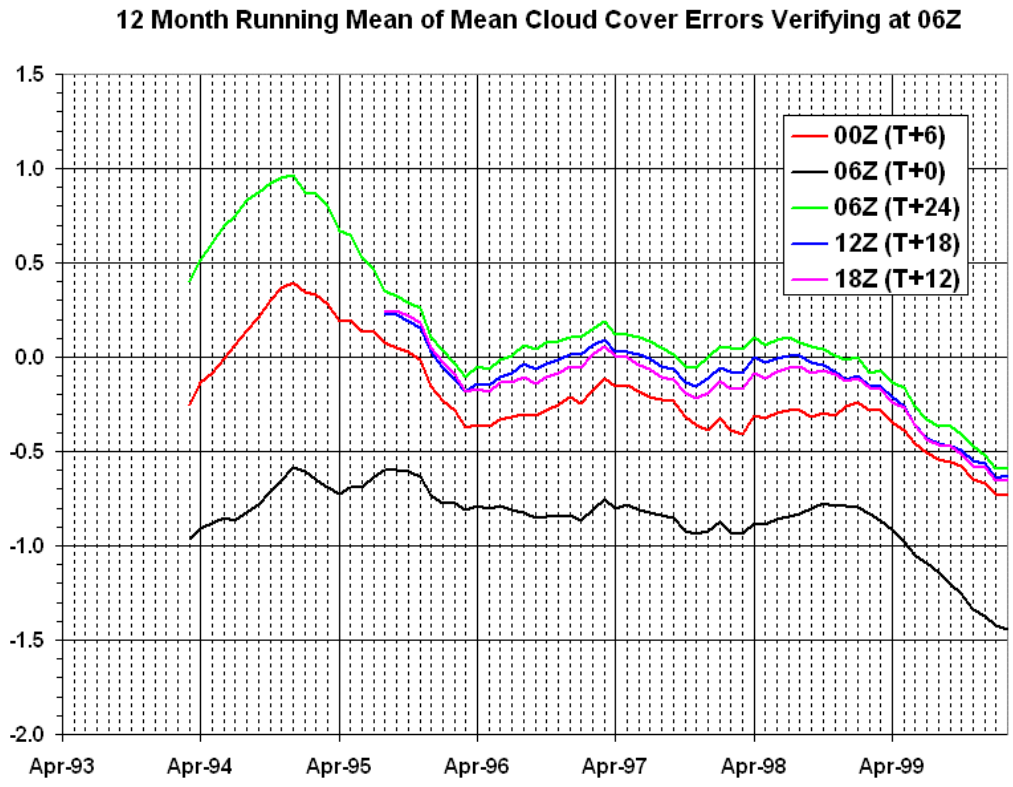

Figure 4.1: Mean cloud cover error in oktas of mesoscale forecasts of different lengths as indicated, all verifying at 06 UTC. Results averaged over ~40 UK stations between April 1993 and the end of 1999.

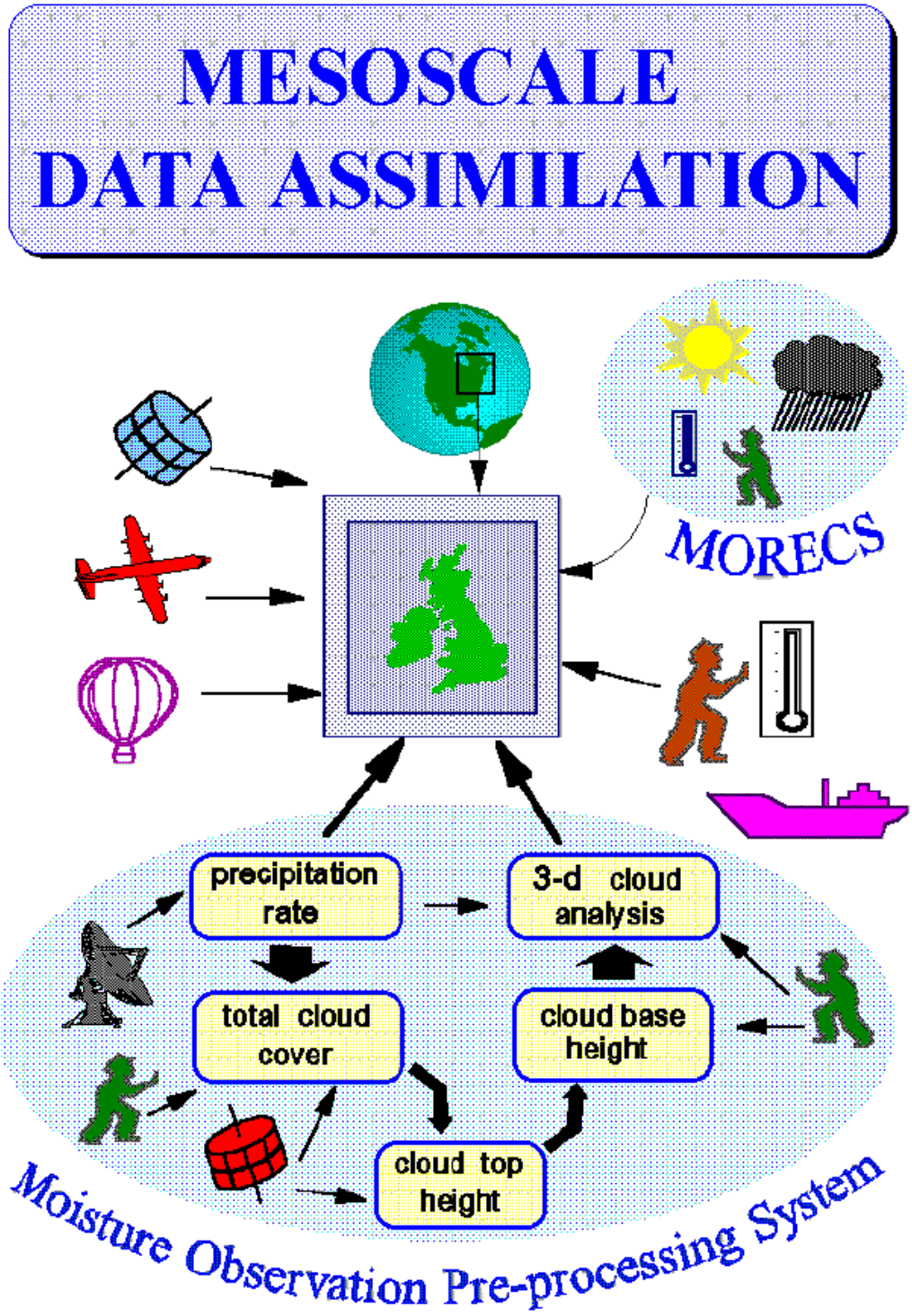

Figure 4.2: The mesoscale assimilation system and data sources. The Moisture Observation Pre-Processing system (MOPS) is effectively an interface with the Nimrod nowcasting system (Golding, 1998), and provides 3-dimensional cloud data and a surface rain rate analysis. (MORECS, the Met Office Rainfall and Evaporation Calculation System, supplies weekly soil moisture data. For its assimilation, see Appendix A of Radcliffe et al., 2001).

4.2 Impact of radiosonde relative humidity

Radiosonde rh data were found to increase the existing negative cloud bias and degrade the screen temperature forecast (compare the blue and red lines on Figures 4.3 and 4.4 below). Earlier work (Lorenc et al., 1996) had already led to operational use of statistically based bias correction for sonde rh near saturation, to give better agreement with surface cloud reports. This study suggested that a 'real-time' bias correction may be needed, in which we compare each sonde with its collocated cloud report. Such a scheme has since been formulated and tested, although it is not yet operational. A preliminary account is given in Sharpe and Macpherson (2001).

Note that these results do not imply in general that radiosonde humidity data are detrimental to the model. They were obtained for an anticyclonic period with little precipitation. Anderson et al. (2000) document the benefit of radiosonde rh data for precipitation forecasts.

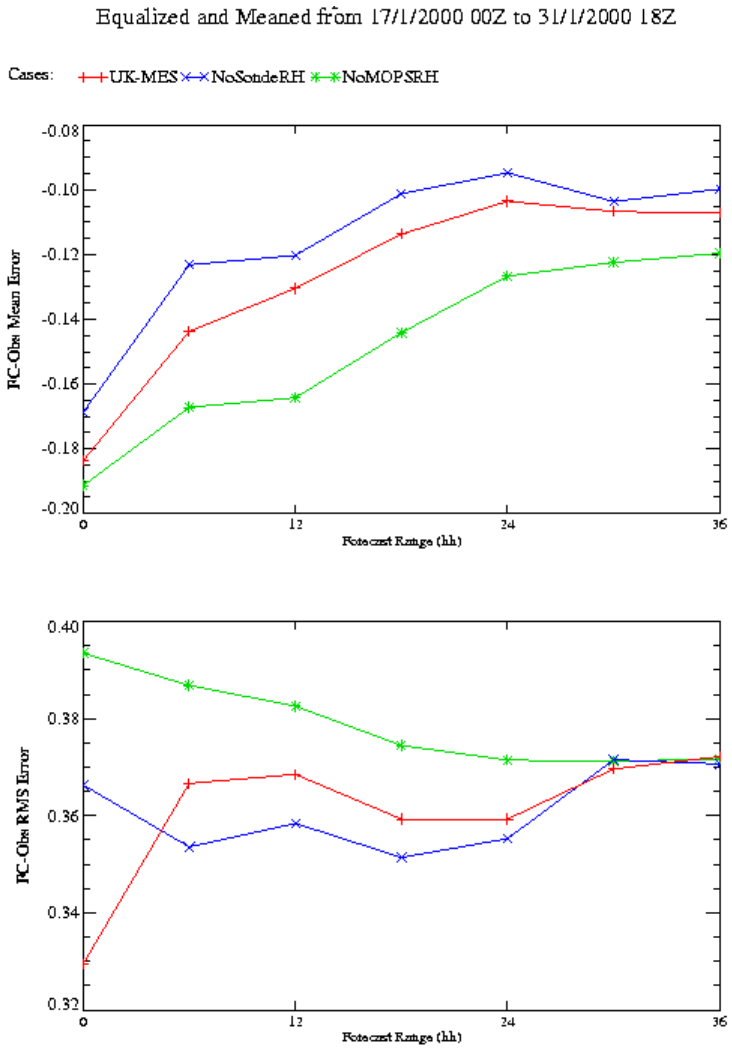

Figure 4.3: Cloud cover verification for period 17th-31st January 2000. Mean error in cloud fraction (top panel) and rms error (lower panel) as a function of forecast time to t+36. Red line for operational control, blue line for NoSondeRH, green line for NoMOPSRH.

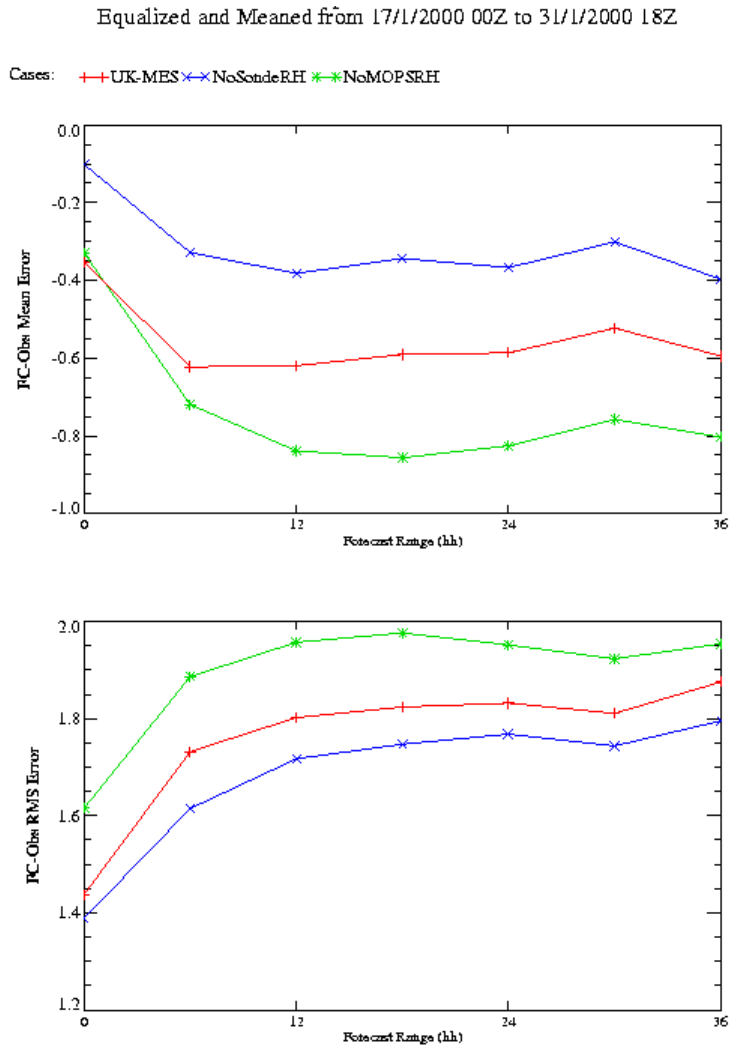

Figure 4.4: Screen temperature verification for period 17th-31st January 2000. Mean error in degrees (top panel) and rms error (lower panel) as a function of forecast time to t+36. Red line for operational control, blue line for NoSondeRH, green line for NoMOPSRH.

4.3 Impact of MOPS cloud data

MOPS cloud data were found to add cloud to both analysis and forecast, in general (compare green and red lines on Figure 4.3), giving a forecast improvement for both cloud cover and screen temperature. However, one period showed MOPS data removing cloud to make the negative bias worse (see green and red lines around 28th January on Figure 4.5, below).

Investigation of this behaviour revealed an explanation in terms of the logic for selection of MOPS cloud data from the full 3-dimensional Nimrod cloud analysis. This led to an algorithm change discussed in detail in the next section.

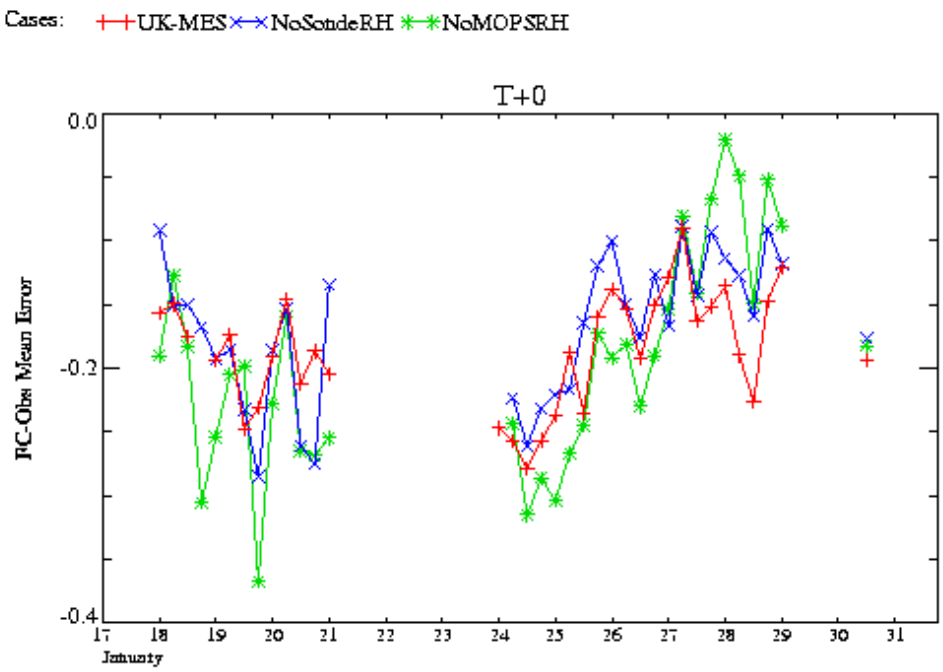

Figure 4.5: Time series of mean error in t+0 fractional total cloud cover throughout period 17th-31st January 2000. Red line for operational control, blue line for NoSondeRH, green line for NoMOPSRH.

4.4 Revision to generation of MOPS cloud data

MOPS data are created by selecting a subset of the full 3-dimensional Nimrod cloud analysis. The selection seeks to identify data which contain new observational information, as opposed to data from the Nimrod first guess advected from the previous hourly analysis. Data are also rejected where they may be vulnerable to inaccurate height assignment of satellite cloud tops, especially for thin cirrus. This led to criteria which allowed MOPS to use data where the satellite had *removed* cirrus diagnosed to be present at the wrong level in the first guess. On the other hand, data were not selected if the satellite implied *addition* of cirrus at some other level, for fear of this level being too inaccurate and leading to forecast errors, perhaps through initiation of spurious convection. This meant an undesirable tendency for MOPS data to remove cloud from the model analysis during assimilation when cirrus was present.

Several desirable features of the MOPS data selection algorithm were proposed:

- disallow use of data at a level seen by satellite when the cloud top is above 5km, even when the data indicate cloud removal relative to Nimrod first guess
- disallow use of data at a level seen by surface reports if it is also above the lowest level seen by satellite and the cloud top is above 5km.
- disallow use of data from levels seen by neither satellite nor surface station

The first deals with the asymmetry in addition/removal of cloud by satellite above 5km.

The second feature was deemed prudent in case use of satellite data above 5km compromised use of some surface reports. The logic in place allowed for the flag 'seen from the surface' to result in selection of the data even if it was also seen by satellite (and with a cloud top greater than 5km). But satellite data could add spurious cloud below the top level seen from the surface, and some points near a surface observation may be affected by the surface report but not so completely as to remove any incorrectly inserted satellite cloud. The one possible exception to this is where clear sky is observed by both surface station and satellite, and the zero cloud fraction data should be used throughout the whole column. (Actually, a threshold of 1 okta is used, below which cloud is reset to zero).

The third condition was considered advisable because such data could lead to cloud removal at any level, irrespective of the cloud top height, or cloud addition if cloud top is below 5km. This data results from the scaling of 3-dimensional cloud profiles to make them consistent with the total cover analysis, but some of the information so derived is not at levels actually observed, so may be unsafe to assimilate. The scaling always makes changes of the same sign throughout the column, whereas it might be better to add cloud at some levels and remove it at others.

The resulting new selection logic is to use cloud data from a level:

IF(seen from surface.AND. .NOT.(seen by satellite)
.OR.(seen by satellite.AND.(cloudtop<5km))
.OR.(total cover<threshold of 1 okta))

A trial of this revised MOPS quality control was run in February 2000, and showed a reduced bias in the model cloud analysis (Figure 4.6). There was little forecast impact, probably because the model is relatively quick to regenerate cirrus that is removed in the analysis. This revised MOPS version was introduced operationally at the end of March 2000. It led immediately to a subsequent improvement of around 0.4okta in the operational t+0 cloud bias for April 2000.

Returning to the increasing cloud cover bias in Figure 4.1, we have demonstrated that the bias at t+0 is sensitive to changes in MOPS data. This begs the question as to what caused the original increase in bias during 1999. The explanation appears to be that the previous version of the Nimrod cloud analysis, with its tendency to remove cirrus, was effectively introduced in January 1999. It was then that an error in Nimrod was corrected that had prevented any influence from satellite imagery, probably since spring 1997.

The trend in bias during the forecast is another matter. This can probably be ascribed to various model physics changes. Further reduction in cloud bias during both analysis and forecast was achieved through a change to the cloud scheme introduced at the beginning of June 2000 along with the 3DVar upgrade. At the end of 2001, cloud biases have returned to spring 1999 levels.

4.5 New Nimrod cloud detection scheme

A further change to the Nimrod cloud analysis became operational in November 2001. The improved cloud detection scheme (Watkin, 2001), involved a major upgrade of the infrared based component. It gave a general improvement in the Nimrod total cloud cover analysis, especially over the sea. Impact on model forecasts from assimilating the new data was not measured.

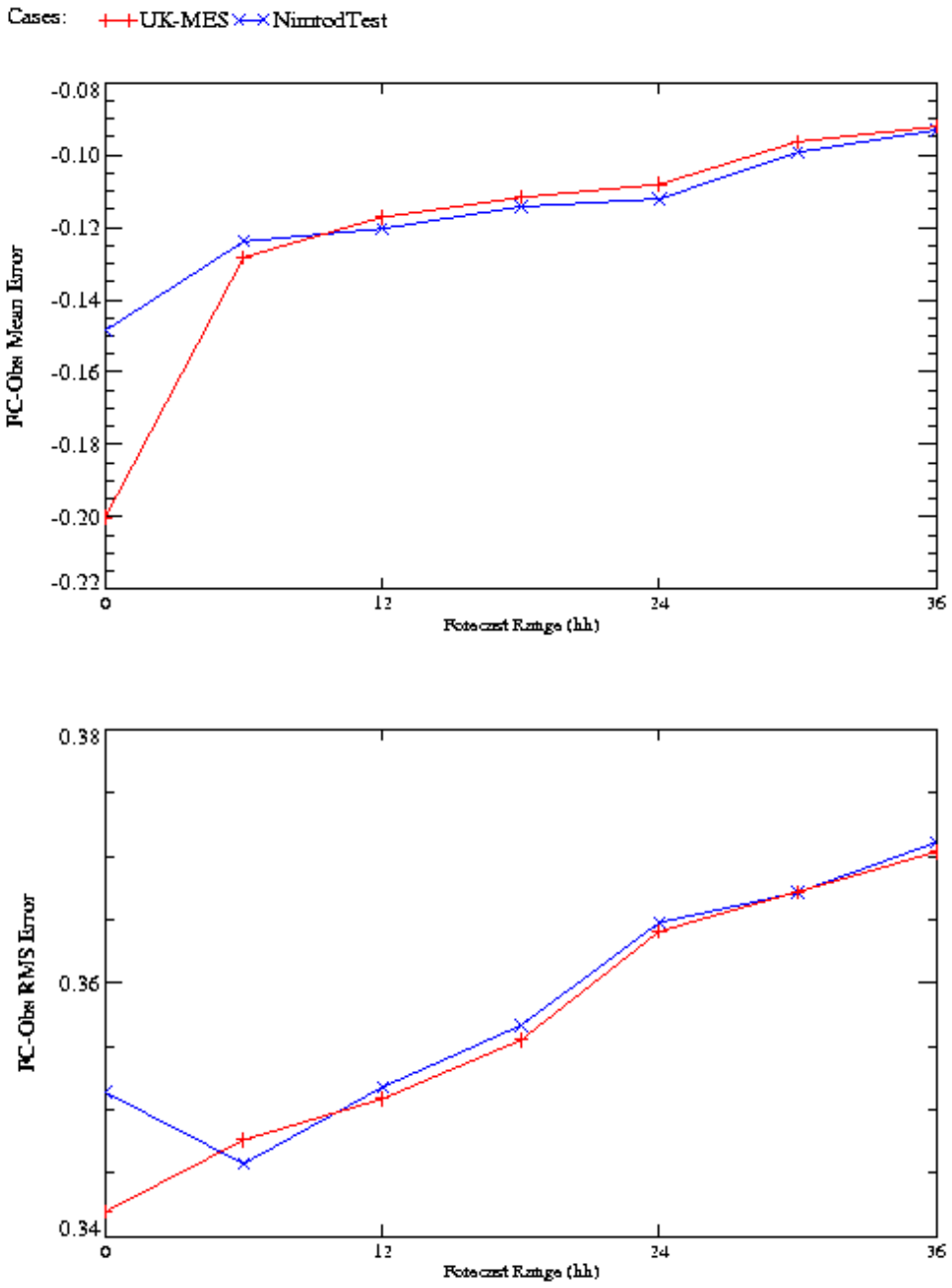

Figure 4.6: Trial of revised data selection algorithm for MOPS cloud analysis. Mean error (top panel) and rms error (lower panel) in fractional cloud cover as a function of forecast time. Results averaged over UK stations for period 12/2/00 to 13/3/00. Red line for operational control, blue line for revised MOPS .

5. Developments in rainfall assimilation

5.1 Motivation for a new radar quality estimate

The original version of the latent heat nudging (LHN) scheme for rain rate assimilation (Jones and Macpherson, 1997) saw little change for four years. Some impact studies were run and data frequency increased from 3-hourly to hourly, work reported by Macpherson (2001).

The treatment by Jones and Macpherson (1997) of the error characteristics of radar based surface precipitation rate estimates was very simple. Relative to the model background field, radar data was assumed to be of uniformly high quality within 100km of the nearest radar and of diminishing quality out to maximum range around 200km.

- One weakness of the implementation of this idea is the assumption of a *fixed* radar network, whereas at any one analysis time, one or more radars may not be contributing to the composite. This means that some of the data assumed to be coming from a nearby radar may actually come from a radar further away, and the Latent Heat Nudging (LHN) scheme may give too much weight to the longer range data. A real-time assessment of data quality is required.

- Secondly, the original system was inflexible with respect to introduction of new radars into the network, requiring a code change in the assimilation to update a list of radar co-ordinates. With the introduction of new European radars into Nimrod, it was timely to make the assimilation code independent of the precise radar network.

- More fundamentally, surface precipitation rate estimates are found to be less accurate when the radar beam is above the freezing level. This is primarily due to the difficulty in extrapolating the observed reflectivity above the freezing level to the reflectivity that would be observed at ground level, which in turn is used to determine the precipitation rate at the ground. Therefore, it is desirable to have a quality measure which takes account of radar beam geometry and freezing level height.

5.2 Formulation of the radar quality estimate

In more detail, the weighting factor, W, in the original Mes LHN scheme is given by:

$$\begin{aligned} W &= W_0 (R - d)/(R - 100) \qquad d > 100\text{km} \\ W &= W_0 \qquad\qquad\qquad\qquad\quad d \leq 100\text{km} \end{aligned} \tag{1}$$

where W_0 is a constant (set to 10), R is the maximum range of the radar, and d is the radar-pixel distance, both in km. W gives the relative weight of the radar precipitation rate to the model precipitation rate when calculating the analysed precipitation rate, towards which the model is forced during LHN.

The new quality measure is given by:

$$W = W_0\alpha \qquad h < h_{fl}$$

$$W = W_0\,\alpha\left(1-\frac{(h-h_{fl})}{4}\right) \qquad h_{fl} < h < (h_{fl} + 4\text{ km})$$

$$W = 0 \qquad h \geq h_{fl} + 4\text{ km} \qquad (2)$$

where W_0 is a constant, h is the beam height, h_{fl} is the height above the freezing level and α is a factor that further reduces the weighting beyond 150 km, given by:

$$\alpha = 1 \qquad R < 150\text{ km}$$
$$\alpha = 1-\left(\frac{R-150}{60}\right) \qquad 150 < R \leq 210\text{ km}. \qquad (3)$$

Equations (2) and (3) essentially say that the radar data quality should decrease once the beam is above the freezing level, falling to zero when the beam is more than 4 km above the freezing level. This should be a reasonable measure of radar quality, since once the beam height is higher than 4km above the freezing level it is difficult to reliably extrapolate to a rainrate at the ground. The factor α means that the data quality also falls off once the range is greater than 150km from the radar. This takes into account the effects of increasing beam width at longer range, which leads to a reduction in data quality.

For a winter case with relatively low freezing level, the new quality estimate produces a weights field like Figure 5.1(a), while a summer case is simulated by a field for constant freezing level of 3500m in Figure 5.1(b). The original weights produced by equations (1) are in Figure 5.1(c). Note the difference between Figures 5.1(a) and 5.1(c) in NW Scotland where the freezing level is low for the winter case.

One qualification of the above formulae is applied, in that we ensure for wet observations, $W \geq 1$. This means we still give significant weight to a wet observation near the edge of the radar area, which could be important if the model were dry there.

5.3. Trial of radar quality estimate

A continuous assimilation 'parallel' to the operational system was run for 3 weeks from 25th June to 16th July 2000, with forecasts from the operational and trial systems verified objectively against Nimrod 6-hourly accumulation analyses, as well as standard surface observations. Precipitation output was also assessed subjectively. A reasonable variety of weather was encountered in the trial period.

5.3.1 Subjective assessment

At least one pair of runs for each day of the parallel trial was assessed, with all 4 pairs of runs compared during the first week. No pair of runs had large enough differences for a preference to be made. Some small differences did occur but these were always minor compared to

differences between the forecasts and actual radar precipitation.

5.3.2 Objective verification

The five variables comprising the UK index were verified against a station list of 'all UK stations'. Results are in the table below. Precipitation scores come from verification against Nimrod accumulations.

	Skill/ETS	
	operational	trial
wind	0.675	0.674
temperature	0.585	0.585
precipitation	0.200	0.198
cloud cover	0.167	0.166
visibility	0.083	0.080
UK index	0.342	0.3405

The trial and operational scores for all variables are very close, and the overall index shows a -0.4% impact, which is small enough to regard as neutral. The breakdown of precipitation impact according to threshold is seen in Figure 5.2. For 3 out of 4 thresholds, the signal is very neutral. The only one for which there seems to be a slight detriment is for 4mm/6 hours, but the occurrence of this threshold was relatively low in the period. Certainly the subjective assessment did not pick up any problems.

Conclusion

In view of its better scientific foundation and technical benefits, a neutral impact from the new quality estimate was judged adequate for implementation, which took place on 26th September 2000.

5.4 Improvements in radar data coverage

On February 21st 2001, a significant milestone for the Mesoscale assimilation system was reached, when we began operational assimilation of an extended area Nimrod radar composite including 6 French radars and one Belgian radar. This was the first improvement in areal coverage of radar data since the LHN scheme was implemented in 1996. The aim of acquiring the extra data is to improve model rainfall forecasts for systems reaching the UK from the south, particularly in summer time. The extended area is shown in Figure 5.3, and with addition of further radars from France is expected to grow early in 2002. Netherlands radars are also due for incorporation. The quality of the extra radar data has been assessed by Gibson (2000) and Gibson (2001).

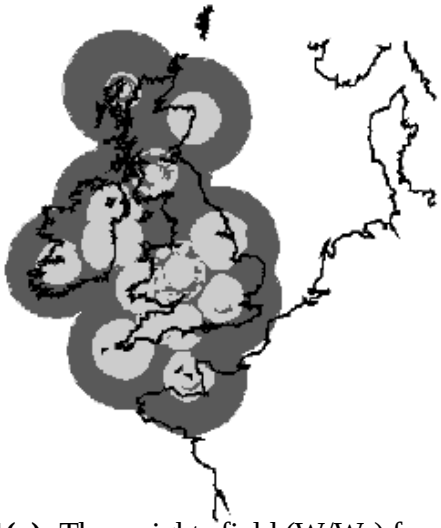

Figure 5.1(a): The weights field (W/W_0) from equations (2) and (3) for a winter case. Light shading for (W/W_0)=1, dark shading for (W/W_0)<1

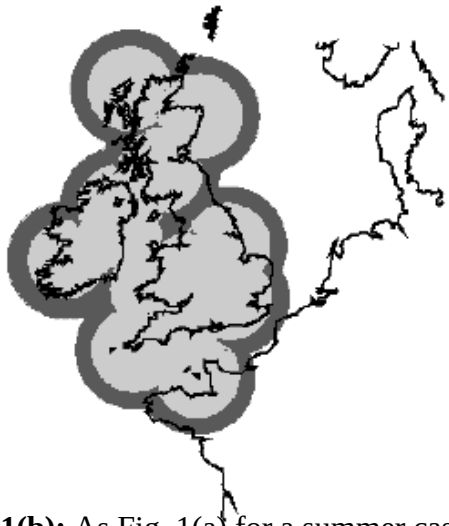

Figure 5.1(b): As Fig. 1(a) for a summer case with uniform freezing level height of 3500m.

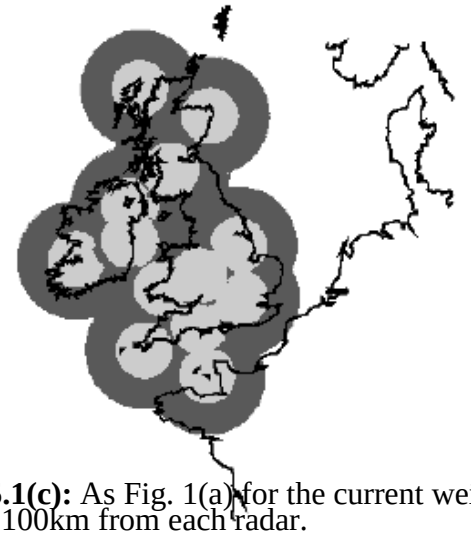

Figure 5.1(c): As Fig. 1(a) for the current weights given by equations 1, for which (W/W_0) =1 up to a range of 100km from each radar.

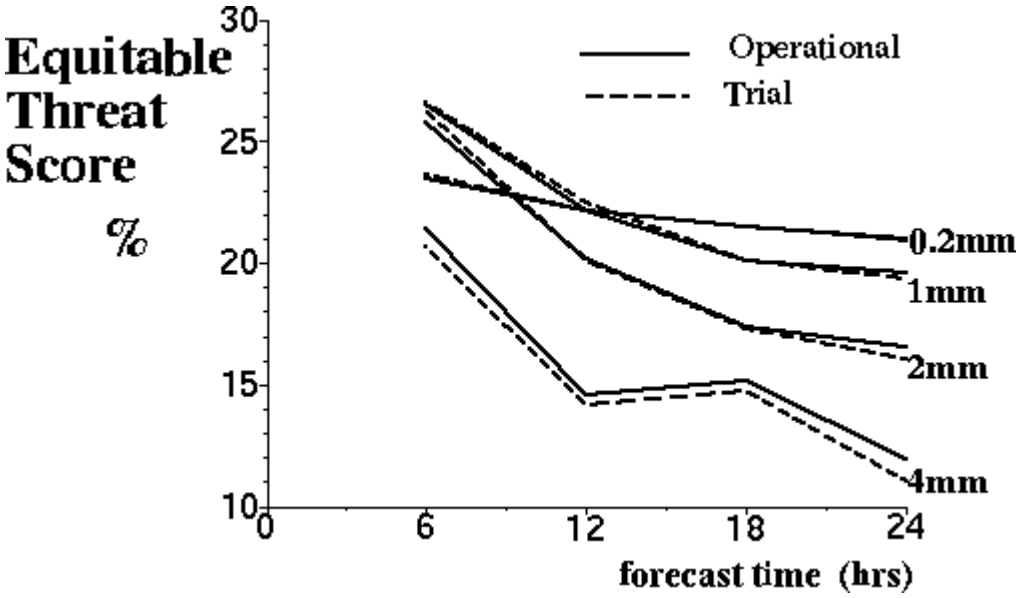

Figure 5.2: Trial of radar quality estimate. Verification of 6-hourly accumulations v Nimrod analyses for 45 forecasts in the period 28th June to 17th July 2000. Equitable Threat Score (%) as a function of forecast time for 4 thresholds. Solid line for Operational, dashed line for Trial.

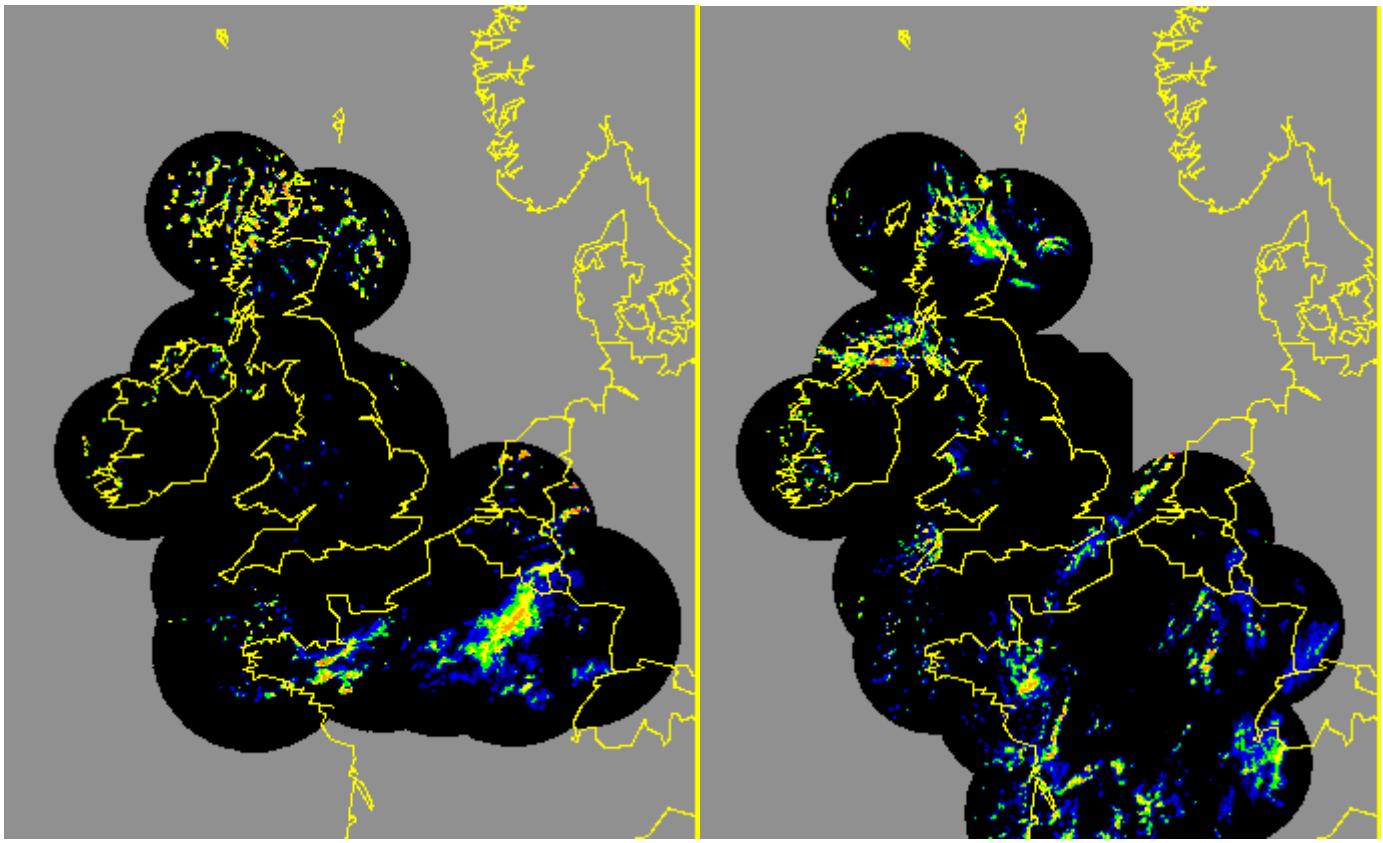

Figure 5.3: extended radar coverage area implemented in February 2001 (left frame). Further extended area of radar coverage due for implementation in early 2002 (right frame). Compare with standard verifying radar frames from UK network in, for example, Figures 3.3 or 2.3.

6. References

Anderson, S.R., Graham, R.J. and M.J. Bader, 2000: The Impact of Observations on Mesoscale Model Forecasts of 3-hourly Rainfall Accumulations. Meteorol. Appl. **7**, 193-203.

Gibson, M. 2000: Report on European radar data for use in Nimrod. Forecasting Research Technical Report No. 326.

Gibson, M. 2001: Report on further European radar data for use in Nimrod. Forecasting Research Technical Report No. 347.

Golding, B.W., 1998: Nimrod: A system for generating very short range forecasts. Meteorol. Appl. **5**, 1-16.

Ingleby, N.B. 2001: The Statistical Structure of Forecast Errors and its Representation in The Met. Office Global 3D-Var. Quart. J. Roy. Met. Soc., **127**, 209-231.

Jones, C.D. and Macpherson, B. 1997: A Latent Heat Nudging scheme for the Assimilation of Precipitation Data into an operational Mesoscale Model. Met Apps, **4**, 269-277

Lorenc A.C., Barker D., Bell R.S., Macpherson B. and Maycock A.J. 1996: On the use of Radiosonde Humidity Observations in mid-Latitude NWP. Meteorol. Atmos. Phys. **60**, 3-17.

Lorenc, A.C., S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne and F.W. Saunders. 2000: The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. Quart. J. Roy. Met. Soc., **126**, 2991-3012.

Macpherson, B. 2001: Operational Experience with Assimilation of Rainfall Data in the Met. Office Mesoscale Model. Meteorol. Atmos. Phys., **76**, 3-8.

Macpherson B., Wright B.W., Hand W.H. and Maycock A.J. 1996: The impact of MOPS moisture data in the UK Meteorological Office data assimilation scheme. Mon. Weath. Rev. **124**, 1746-1766.

Radcliffe, J., Chalcraft, B.V., Wilson, C.A., Macpherson, B., Wilson, D.R., and P.A.Clark. 1998: Report on pre-operational trials of the new microphysics scheme in the mesoscale model. Forecasting Research Technical Report No. 259.

Radcliffe, J., Chalcraft, B.V., Wilson, C.A., Macpherson, B., Jones, C.P., Cox, P.M. and M.J. Best. 2001: Pre-operational testing of the new boundary layer and MOSES-I surface exchange schemes in the UK Mesoscale Model. Forecasting Research Technical Report No. 343.

Sharpe, M. and Macpherson, B. 2001: Developments in the correction of radiosonde relative humidity biases at the Met Office. Preprints 18th AMS Conference on Weather Analysis and Forecasting and 14th Conference on Numerical Weather Prediction, Florida, 30th July-2nd August, pp 250-252.

Watkin, S. 2001: An improved cloud detection scheme for Nimrod. Forecasting Research Technical Report No. 371.